

TUTORIAL

Evaluation of covariate effects using forest plots and introduction to the *coveffectsplot* R package

Jean-Francois Marier¹  | Nathan Teuscher¹  | Mohamad-Samer Mouksassi^{1,2,3} 

¹Certara Strategic Consulting,
Princeton, New Jersey, USA

²Université de Montréal, Montreal,
Québec, Canada

³Lebanese American University, Beirut,
Lebanon

Correspondence

Mohamad-Samer Mouksassi, Certara
Strategic Consulting, Princeton, NJ
08540, USA.

Email: samer.mouksassi@certara.com

Funding information

No funding was received for this work

Abstract

The current tutorial describes why forest plots are needed for an effective communication of covariates effects, how they are constructed, and how they should be presented. Simulation-based methodologies allowing the user to evaluate the marginal impact of changing one covariate at a time or by considering the joint effects of correlated covariates are introduced along with graphical tools for an optimal assessment of the covariate effects. The R package *coveffectsplot* and an associated R Shiny application are provided to facilitate the design and construction of forest plots for the visualization of covariate effects. All codes and materials are available on a public Github repository.

Study Highlights**WHAT IS THE CURRENT KNOWLEDGE ON THE TOPIC?**

Various methodologies have been used to assess covariate effects in pharmacometrics projects. There is currently a need to develop software framework and workflows for producing and interpreting covariate effects/range of possible exposures.

WHAT QUESTION DID THIS STUDY ADDRESS?

A set of workflows for assessing covariate effects are proposed with the goal of clearly communicating the underlying methodology used for simulations along with an R Shiny application that can facilitate the presentation of results and inform potential decisions regarding dose adjustments in subpopulations of interest.

WHAT DOES THIS STUDY ADDS TO OUR KNOWLEDGE?

A simulation-based methodology allowing the user to evaluate the impact of marginal effects changing one covariate at a time or by taking into consideration the joint effects of distributions of correlated covariates is introduced along with an associated R package that include an R Shiny application to streamline the design of plots that help with the communication of results.

HOW MIGHT THIS CHANGE DRUG DISCOVERY, DEVELOPMENT, AND/OR THERAPEUTICS?

The proposed software framework is expected to standardize and streamline the assessment and interpretation of covariate effect which may in turn accelerate the execution of projects due to better communication within drug development teams and with regulatory agencies.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) 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.

© 2022 The Authors. *CPT: Pharmacometrics & Systems Pharmacology* published by Wiley Periodicals LLC on behalf of American Society for Clinical Pharmacology and Therapeutics.

INTRODUCTION

An important focus of pharmacometrics is to understand variability in pharmacokinetic (PK), pharmacodynamic (PD), and clinical response. Covariate analyses are performed to identify sources of variability in model parameters to support drug development decisions, such as dose adjustments in subpopulations of interest to optimize efficacy and safety of treatments, or to allow prediction (interpolation or extrapolation) of outcomes under new experimental conditions, such as untested doses, regimens, populations, or disease factors.

Effective communication of quantitative findings was covered in a previous tutorial with an emphasis on understanding the targeted audience, speaking an understandable language, and provided clear data that may support drug development decisions.¹ For an in-depth review of effective visual communication, refer to Vandemeulebroecke et al.,² in which the following three principles of successful communication were presented: a purpose, a clear data presentation, and an obvious message.

The current tutorial describes the use of forest plots to communicate the impact of covariates in pharmacometrics model and the preparation of those plots using the R package *coveffectsplot*.³ An R Shiny application with a user-friendly interface is introduced to facilitate the construction of forest plots and to ease the potentially steep learning curve of producing quality graphics. Examples involving the impact of changing one covariate at a time or by considering full distributions of correlated covariates are presented.

WHAT ARE FOREST PLOTS

Forest plots were initially used to display results of multiple clinical studies on a single plot, and subsequently used to display results of cross-study meta-analyses.^{4,5} In general, a forest plot contains a set of variables of interest stacked vertically, and, for each variable, a point estimate and associated intervals are displayed. The point estimate is relative to a reference value such that the relative importance of each variable of interest can be compared. An example of a simple forest plot is presented in [Figure 1](#) whereby the variables of interest (weight and sex) are on the left side of the plot, each occupying a separate panel delimited by a gray rectangle with the covariate name in the middle. Each row within the panel represents a different covariate value defined on the left axis tick marks (e.g., 128 kg, 98 kg, male, etc.). The blue circles represent the point estimates for the effects of the covariate of interest and here the associated horizontal lines are the 95% confidence intervals (CIs), relative to the reference subject.

In the current example, the reference is a female subject with a body weight of 85 kg and the point estimate (1 by definition). The dashed vertical line represents the ratio of 1.0 (the reference) and the gray area represents an area of clinical significance (e.g., 0.8–1.25). The dashed line and gray area extend to all panels to help in setting a common visual reference. The beauty of a forest plot is the ease with which a reader can understand the data. The reader can quickly realize that extreme body weight values (i.e., 128 and 56 kg) have an important effect on the area under the curve (AUC), whereas sex has a minor effect. Throughout the paper we will present several examples of forest plots where the presentation becomes more elaborate. For example, adding a table with numerical values, showing multiple parameters at the same time (separated by facet or overlaid), and adding more text and explanations in the legends, titles, and captions.

WHY FOREST PLOTS ARE NEEDED

Although forest plots were historically used to present results from a cross-study meta-analysis, its application to illustrate the effect was mentioned in the Population PK Guidance.⁶ Menon-Andersen et al.⁷ suggested that the presentation of PK parameters using forest plots is essential for supporting drug dosage decisions. The authors included examples of clinical pharmacology studies by providing PK parameters computed using noncompartmental analysis methods with a specific focus on drug–drug interactions and special populations (i.e., elderly population and renal or liver impairment). More recently, Xu et al.⁸ proposed a methodology that can be applied in the context of a population PK analysis using a full covariate modeling approach, where simultaneous CIs are studied and forest plots of covariates effects on model parameters are presented.

In particular, forest plots can provide the impact of many variables in a single display so that readers can quickly determine the relative impact of each variable of interest relative to the reference subject. Forest plots are essential for translating quantitative findings to interdisciplinary team members and ultimately support drug development and/or regulatory decisions. Forest plots can be created from post hoc parameters, observed data, or simulated results. Parameter uncertainty can be included alone or in combination with between-subject variability (BSV) using a full variance–covariance matrix, univariate standard errors, or nonparametric distributions. The selection of variable values of interest can influence forest plots (e.g., minimum, maximum, 5th, 10th, 90th, and 95th percentiles). Thus, there is a need to develop a software framework to ease the production of the visual displays,

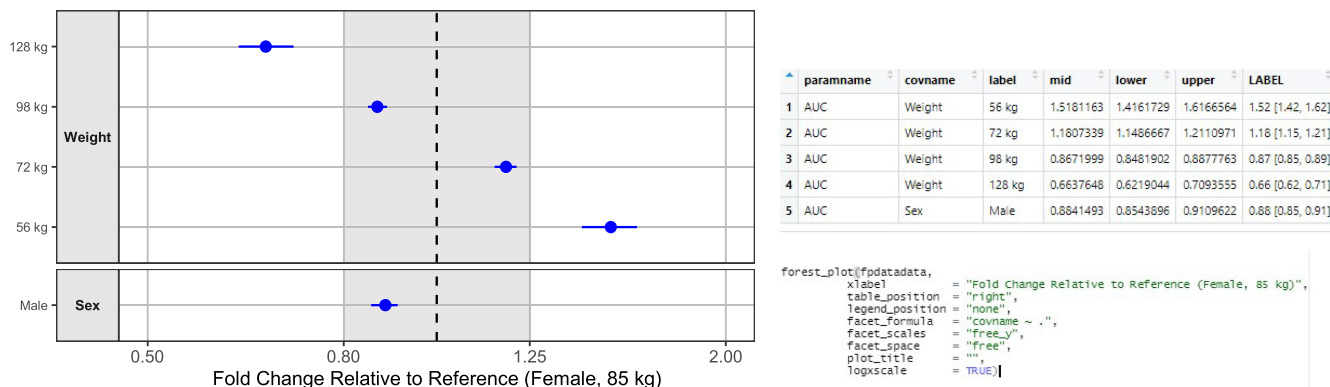


FIGURE 1 Example of a covariate effect plot using the forest plot format. AUC, area under the curve.

which in turn can facilitate the assessment and interpretation of covariate effects plots.

WHEN FOREST PLOTS SHOULD BE PRESENTED

Forest plots are most effective when the measure of interest in the plot is a clinically relevant parameter (e.g., maximum plasma concentration [C_{max}], AUC, and clinical response) rather than a structural component of the mathematical model (e.g., clearance [CL], volume of distribution, and absorption rate constant). Forest plots should be presented as part of the simulation-based assessments of a pharmacometrics model. A common use-case is to create a forest plot to visualize and interpret the effect of covariates on exposure parameters once a final pharmacometric model with covariates is available. The simulations involved for the dataset construction as well as the creation of the forest plots itself can be time-consuming. As a result, the *coveffectsplot* package also includes complete R code workflows that help to standardize and to streamline the assessment and interpretation of (1) covariate effects and (2) range of possible exposures. This, in turn, can accelerate the execution and reporting of pharmacometrics projects.

HOW FOREST PLOTS ARE CONSTRUCTED

Simulation-based workflows are proposed in the current tutorial based on uncertainty of fixed-effect parameters, covariate distribution, and BSV. A graphical presentation of the impact of uncertainty (blue), covariate distribution, such as body weight (green) and BSV (red) on a PK parameter is presented in Figure 2. The contribution of uncertainty, covariate distribution, and BSV on individual clearance estimates are presented in the lower left panel, whereas the standardized clearance relative to a reference

subject is presented on the lower right panel. The current tutorial covers in more detail how forest plots can be derived by implementing the above information as part of simulation workflows. Consistent with the population PK guidance for industry,⁶ two simulation workflows are described in the current tutorial to create forest plots:

1. Simulations based on uncertainty of fixed-effect parameters (example 1).
2. Simulations based on uncertainty of fixed-effect parameters and BSV to predict drug exposure in a future population (example 2) and in the context of a pediatric extrapolation (example 3).

The paper is organized first by presenting three example applications, a brief discussion on tornado plots, and finally a more detailed section on the *coveffectsplot* R package usage as a communication tool. The Supplementary Materials include all the R code with the full details on the parameters used in the simulation, the computation of the exposure parameters and all data processing and plot codes to fully reproduce the presented results.

EXAMPLE 1: FOREST PLOTS - SIMULATION APPROACH INCLUDING UNCERTAINTY OF FIXED-EFFECT PARAMETERS

As stated in the population PK Guidance for Industry,⁶ the uncertainty in fixed-effect parameter estimates is accounted for in the simulations when the goal is to illustrate the probability of the typical subject's drug exposure to reach or stay above a specific cutoff point or if one wishes to illustrate the effect of covariates. These types of simulations can also be used to evaluate the performance of new dosing regimens for testing in future trials or to graphically illustrate the effect of parameter precision on the simulated PK profiles. Presenting the

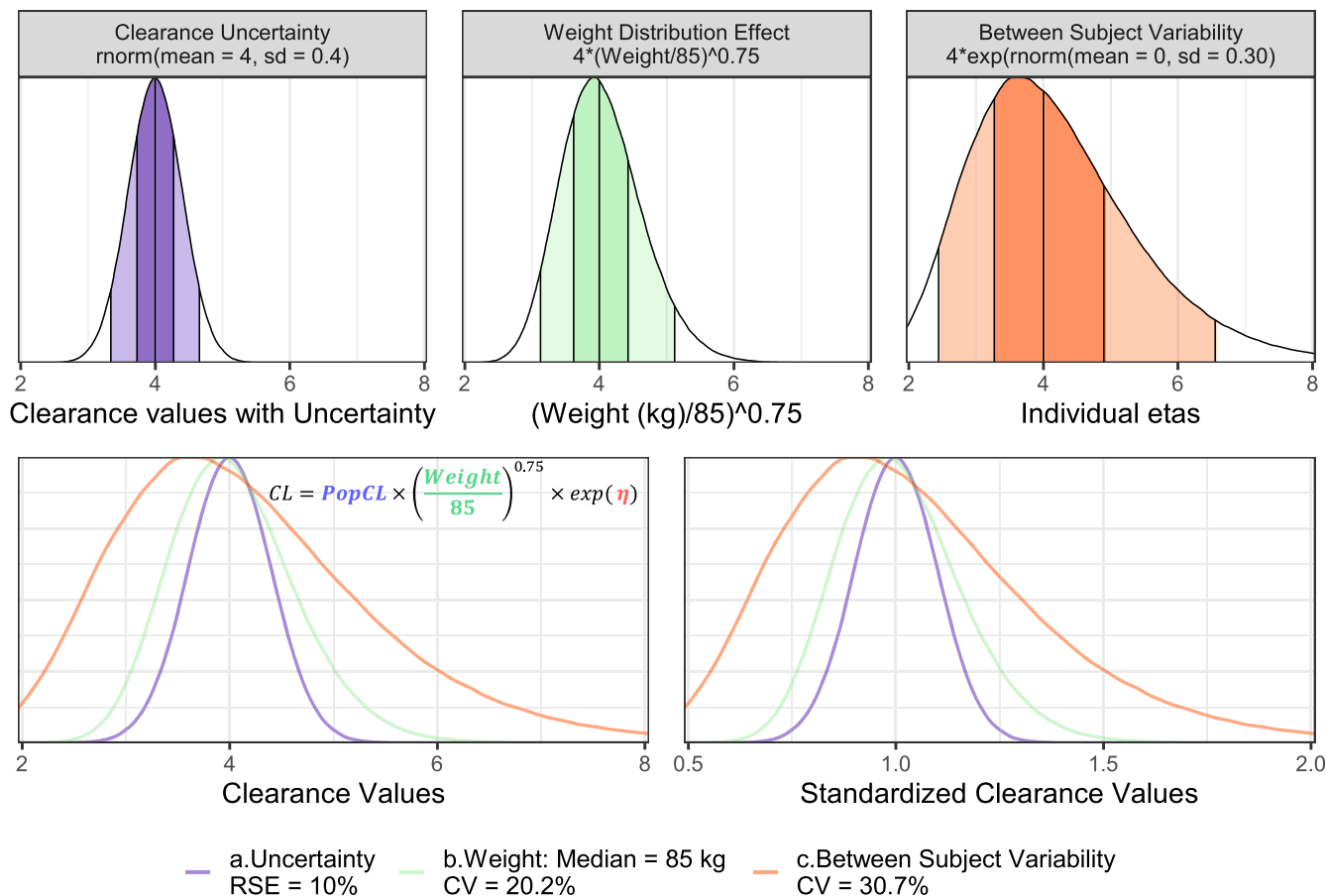


FIGURE 2 Illustrating the impact of uncertainty, covariate distribution, and between-subject variability. CL, clearance; RSE, relative standard error.

computed effects as forest plots facilitates the interpretation of the relative importance of covariates on exposure.

Briefly, this approach consists in changing one covariate at a time while fixing others at their reference values and simulate concentration–time profiles with uncertainty on PK parameters. This approach allows the assessment of a single covariate at a time, which has traditionally been referred as “marginal effects” in statistics.^{9,10} For continuous covariates, the effects at specific values of interest must be computed. A common practice is to use the 5th, 25th, 50th (median used as reference), 75th, and 95th percentiles from the observed covariate distribution. In some cases, the minimum or maximum values of covariates may be considered, although caution should be made regarding potential outliers or aberrant values. For categorical covariates, all categories available in the dataset are accounted for.

Care should be taken, as this approach may generate unrealistic combination of covariates (e.g., when body weight and age in pediatric population are varied independently). The user should keep in mind that the primary purpose of it is to understand the relative importance of

the marginal effects of covariates on the simulated exposure parameter.

A simulations-based workflow for assessing the impact of covariates on drug exposure by performing simulations with uncertainty of fixed-effect parameters only and changing one covariate at a time is presented in [Figure 3](#) (example 1). A two-compartment PK model with a first-order absorption defined with ordinary differential equations was used. The model included the effect of albumin, sex, and body weight on the apparent clearance, whereas body weight influenced the apparent volume of distribution. Covariates were centered by median values of body weight (85 kg) and baseline albumin levels (45 g/L) and using the female category as the reference value. Parameter uncertainty for this model was represented by a variance–covariance matrix that included 15% relative standard errors on all parameters and correlations of 0.2 between each pair of parameters. The proposed workflow is split into the following steps:

Step 1: A template dataset is constructed by varying one covariate at a time, while keeping other covariates consistent with the reference. The reference values are usually chosen as part of the modeling exercise,

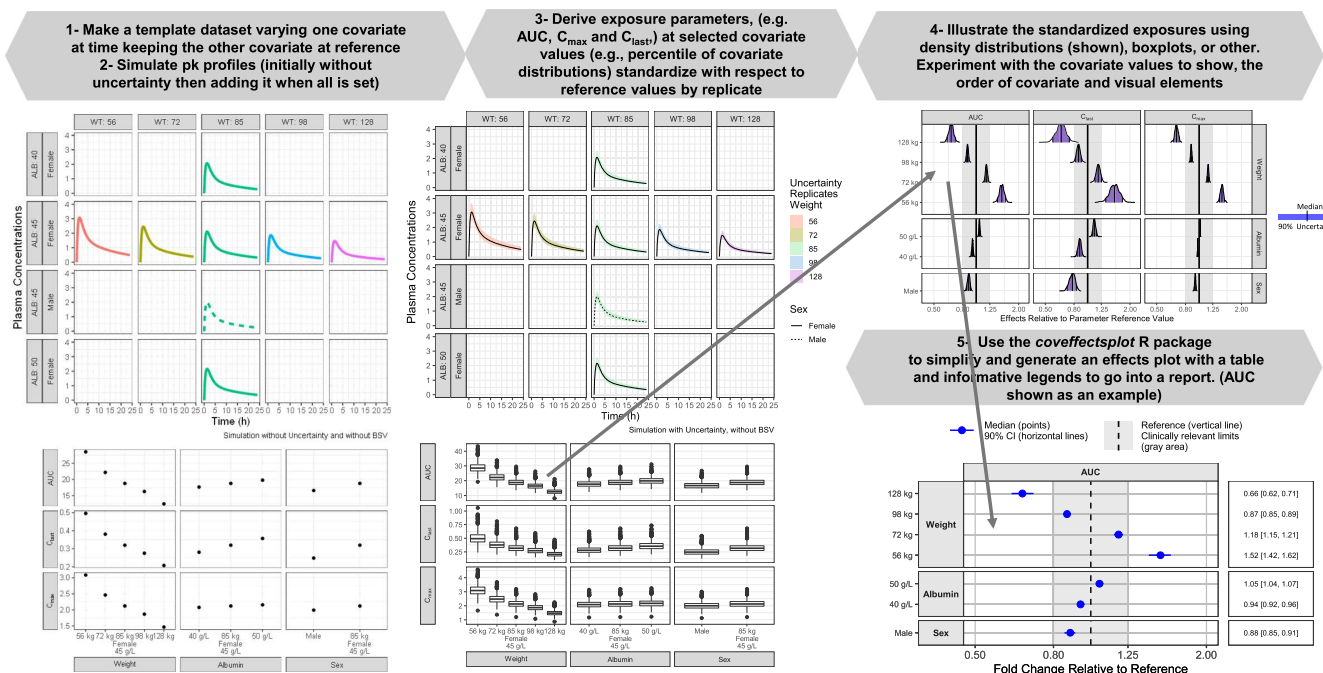


FIGURE 3 Example 1: Simulations based on uncertainty of fixed-effect parameters only. AUC, area under the curve; C_{last} , last measurable concentration; C_{max} , maximum plasma concentration; PK, pharmacokinetic.

however, the user can change the reference values or categories before or after the simulation to better reflect the target population of interest. The template dataset is expected to include all the needed combinations of covariate values. For this example, the following covariate values were used:

1. Weight values at the 5th, 25th, 50th, 75th, and 95th percentiles (56, 72, 85 [reference], 98, and 128 kg, respectively).
2. Albumin values at 40, 45 (reference), and 50 g/L.
3. Sex categories: female (reference) and male.

This would result in having eight unique combinations of covariate values in the dataset template.

Step 2: A preliminary simulation is performed without uncertainty to verify that the model was coded properly and that results are as expected. Concentration–time profiles are simulated for all combinations of covariates generated in step 1 and exposure parameters of interest are derived (e.g., AUC, last measurable concentration, and C_{max}).

Step 3: Concentration–time profiles are re-simulated this time including uncertainty on fixed-effect parameters. The fixed-effect parameters are sampled from the uncertainty distribution (i.e., variance–covariance matrix in this example) and the process is repeated multiple times (e.g., 1000 replicates). A total of 1000 times eight combinations would require simulating 8000 PK profiles. Exposure parameters are then computed on each profile for each

subject. A graphical display (e.g., boxplots) or descriptive statistics summaries of the computed exposures at this stage can be useful as an initial view of the simulation results. Standardized exposure parameters relative to reference patient values in each replicate are then computed. This will result in having a common value of one for all exposure parameters and enable the opportunity to appreciate the distributions in term of ratio ranges.

Step 4: The standardized parameters are illustrated using density. In this example, the 90% ranges of the distributions are shaded in blue. These density plots can be simplified to improve the communication of results using forest plots as explained in the next step.

Step 5: The presentation of the distributions of uncertainty on fixed-effect parameters is simplified using point estimates with intervals. Briefly, the 90% intervals are computed using the 5% and 95% quantiles and a dataset that is compatible with the *coveffectsplot* package is produced. At this stage the main focus is to optimize the communication of results based on the various visual elements of the plot:

1. Which exposure parameters to present: Presenting one or more exposure parameter together or separately? Here, only AUC is kept as a first simple example.
2. Order of presentation of covariates: Weight on top (widest range), albumin next, and sex last. We can notice that when weight increases AUC decreases, whereas when albumin increases AUC increases.

There should be a common logic and rationale applied to all covariates.

3. Order of presentation of covariates values: Make sure the numerical order is respected (e.g., 56, 72, 98, and 128 kg). Do we want to reverse the order for albumin given that the relationship is inverse?
4. A side table is included to enable the reader to know the exact values of the presented data. The user can choose to present the table on the right or below the forest plot.
5. The text for the legends to describe and inform the reader about what are the presented ranges (90% CI) and what are the reference areas and reference lines.

The *coveffectsplot* package enables the user to control every aspect of the presented intervals using R code or using a point and click interface via the included R Shiny app launched with the “*run_interactiveforestplot()*” command. The goal is to help the user to interactively design the plot by quickly trying multiple possible ways to present the data. Many iterations are needed before settling into a good plot and then the user can get the plot code from the app for full reproducibility.

EXAMPLE 2: FOREST PLOTS - SIMULATION APPROACH INCLUDING BETWEEN-SUBJECT VARIABILITY

If the purpose is to predict the range of individual predictions of concentrations or individual exposure in a future population or according to different experimental conditions, then BSV should be added to the simulation framework (in addition to the uncertainty on fixed-effect and random-effects parameters). Simulations that account for BSV and include covariate effects should be conducted in a population with realistic demographic variables and taking into account the correlation between random effects to avoid unrealistic parameter combinations in subjects. Additional details are provided in the population PK Guidance for Industry.⁶ Demographic variables can be obtained from databases or generated by resampling with replacement of the individuals in the original study or by sampling the covariate distributions and their correlations in the target population.

A simulations-based workflow for assessing the range of possible drug exposures and the joint impact of covariates on drug exposures in a future population is presented in Figure 4 (example 2). This example uses the same population PK model as in example 1. The steps are modified to include reading in a database of available covariates rather than varying one covariate at a time and by including BSV and its uncertainty.

Step 1: A database of virtual population of subjects' covariates distributions ($N = 2000$) is read and fed into the simulation. A pairs plot of the covariate distributions is presented to better visualize correlations and frequencies. In this case, PK profiles at specific covariate values can no longer be presented as now we have joint effects of the full distributions of the covariates. For example, the median and 90% prediction interval using quantile splits of the continuous covariates (e.g., below/above median, quartiles, or tertiles) can be presented. Even though we split by quartiles of weight, this is rather showing the joint effects of all covariates and BSV when we look from the weight angle and so on. Steps 2, 3, and 4 remain the same as in example 1 in Appendix S1 with the only difference being on representing results using splits/ranges of covariate distribution rather than at specific value. In this case, the number of simulated profiles in each replicate is 2000 (joint covariate distributions) times 1000 replicates for a total of 2,000,000 PK profiles. This requires significantly more computing time than example 1. Instead of a reference value, we represent the whole population as all subjects which represent the full range of possible exposures when we simulate with BSV and uncertainty.

In step 5, standardized PK parameters are summarized, and the 90% intervals are presented using a forest plot. In this example, we show how multiple exposure parameters can be shown side by side with a table below it instead of on the right.

Based on the results presented in Figure 3 (simulations with uncertainty on fixed-effect parameters only) and Figure 4 we can see that the body weight was the main covariate affecting the AUC and that the range of possible exposures is mostly affected by what quartile of weight you are in. To comment on the need for dose adjustments based on body weight we will need additional consideration regarding the safety and efficacy of the drug.

EXAMPLE 3: FOREST PLOTS - SIMULATION APPROACH INCLUDING BETWEEN-SUBJECT VARIABILITY IN THE CONTEXT OF A PEDIATRIC EXTRAPOLATION

In this example, a population PK model derived from an adult population is used for extrapolation in a pediatric population. The population PK model included an allometric function accounting for the effect of body weight on clearance and volume parameters with exponents of 0.75 and 1, respectively. A virtual population of pediatric patients with weight/age/sex matched information was derived in step 1. The Centers for Disease Control and

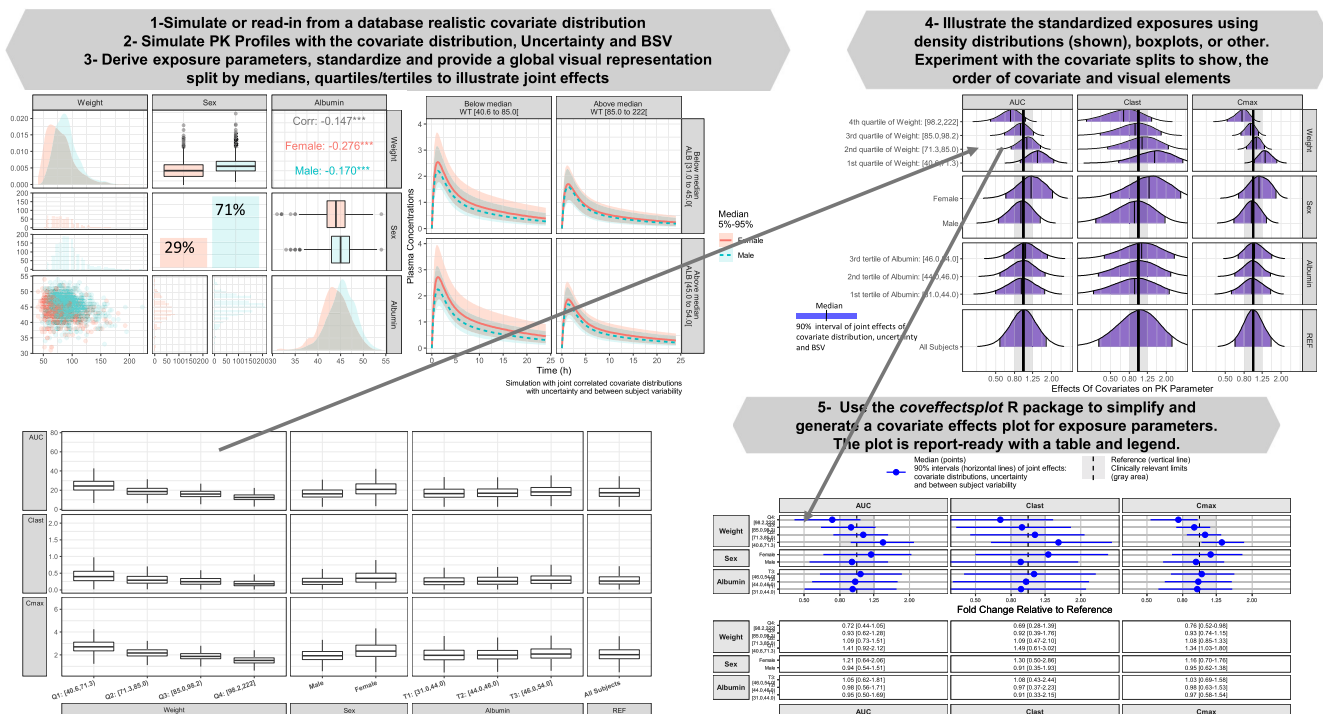


FIGURE 4 Example 2: Simulations based on uncertainty, covariates distributions and between-subjects variability. AUC, area under the curve; BSV, between-subject variability; C_{1ast} , last measurable concentration; C_{max} , maximum plasma concentration; PK, pharmacokinetic.

Prevention (CDC) provides growth chart parameters at specific ages for boys and girls.¹¹ The “*rBCCG*” function from the “*gamlss.dist*” R package is a random number generator for the Box–Cox–Cole–Green distribution. It is useful in our case here, because the CDC provides the values of the fitted parameters (L, M, and S) of this distribution over age and this dataset is included in the *coveffectsplot* package as the *wtage* data object. We can simply plug the provided LMS values into “*rBCCG*” and simulate a virtual population of pediatric patients which takes into account correlation among age, body weight, and sex.¹² The covariate distribution (bottom left panel), along with the resulting weight/age/sex distribution effects plus BSV on exposure parameters are presented as a function of weight (top right) and age (bottom right) in Figure 5.

The complete step-by-step workflow is presented in Figure 6. Briefly, steps 1 to 5 are similar to those presented in example 2, with the exception that a simulation based on a realistic distribution of covariate values for extrapolation purposes was used. Another possible consideration is illustrated in steps 3 and 4 whereby the PK profiles can be presented by a joint two-covariate quantile split (body weight quartiles and age median). This can generate splits with an unequal number of subjects reflecting that some combination will be more frequent than others. Here, we represent AUC and C_{max} versus age and weight and a smoothed quantile regression band to help in understanding how much percent of the population is within

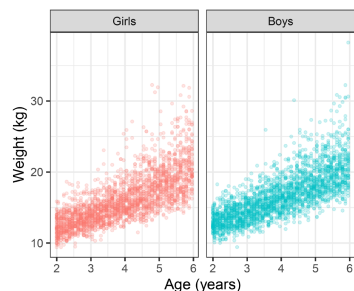
reference limits versus age and versus weight. This can inform whether weight-based or age-based dosing should be used. At step 4, instead of density plots, we show horizontal boxplots as an example, and at step 5, we illustrate how multiple exposure parameters can be represented in the same panel with different symbols. For specific applications, simulations and splits for specific rare populations may be generated (e.g., the malnourished children below 5% percentiles at any age). Overall, the proposed approach is flexible and can be adapted to accommodate complex population splits based on one or multiple combination of factors that may be of clinical importance.

OTHER VISUAL REPRESENTATION METHODS (TORNADO PLOTS)

Tornado plots are used for assessing a change in parameter according to the range of the observed covariate. The low and high values of the covariate affecting the model parameters are used to determine the extent of change in the model parameter relative to the typical value. For example, a population with body weight values ranging from 50 to 90 kg with an allometric model on clearance (e.g., body weight/ $70 \text{ kg}^{0.75}$) is expected to result in clearance values 12% lower (ratio of 0.78) and 21% higher (ratio of 1.21) relative to a typical 70-kg subject, respectively. In tornado plots, covariates are listed vertically, and on each

$$CL = PopCL \times \left(\frac{Weight}{15.8}\right)^{0.75} \times exp(\eta)$$

$$V = PopV \times \left(\frac{Weight}{15.8}\right)^1 \times exp(\eta)$$



The simulated weight/age/sex distribution is fed into the simulation

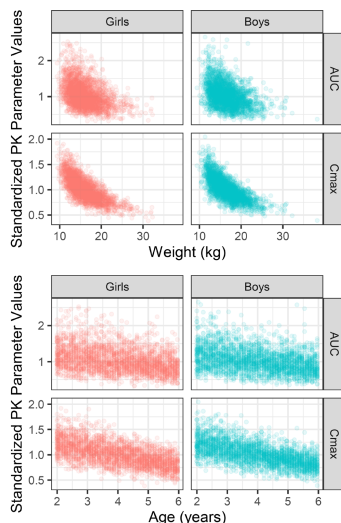


FIGURE 5 Example 3: Simulation of a realistic virtual population of pediatric subjects. AUC, area under the curve; CL, clearance; C_{max} , maximum plasma concentration; PK, pharmacokinetic; PopCL, population clearance; PopV, population volume of distribution.

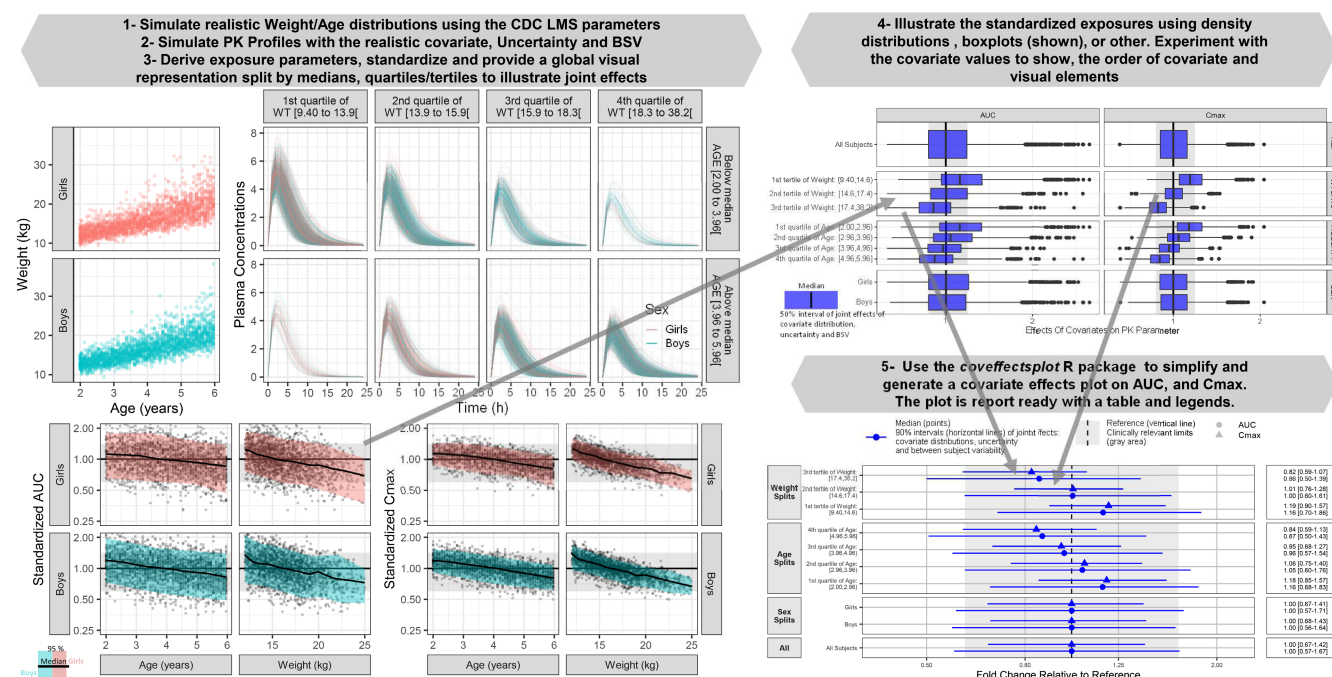


FIGURE 6 Example 3: Assessing ranges of possible exposures by including distributions of covariates, BSV, and uncertainty for a pediatric extrapolation application for 2–6 year old children. AUC, area under the curve; BSV, between-subject variability; CDC, Centers for Disease Control and Prevention; C_{max} , maximum plasma concentration; PK, pharmacokinetic.

row, a bar joining the effects of the low and high covariate value (e.g., 0.78–1.21 as per the above example) is presented. Traditionally, uncertainty is not shown, and rows (covariates) are ordered so that the largest bar appears at the top of the chart, the second largest appears second from the top, and so on.

Depending on how the tornado plots underlying data was generated it can present the marginal or joint effects of multiple covariates with individual post hoc or simulated random effects, with or without uncertainty. Figure S1 presents an example of a tornado plot side by side with a forest plot showing the exact same data. Tornado plots and forest plots are general graphical

techniques that can be used for various applications. In order to correctly interpret their meaning, a detailed documentation on how the shown bars or intervals were computed is needed.

THE COVEFFECTSPLOT R PACKAGE

The *coveffectsplot* R package aims to facilitate the communication of covariate effects by enabling the generation of clear graphical displays using interactivity and flexibility. It helps to layout one or multiple covariate effects on one

or more model parameters with point estimates and intervals presented in a tabular format on the side for completeness. Informative legends and annotation showing areas of interest and or reference lines can be easily edited and included.

The main package function *forest_plot* expects a dataset with a minimum of seven columns with the following exact case-sensitive names:

- paramname*: A parameter name (e.g., AUC and C_{max}).
- covname*: Covariate, reference or splits name (e.g., body weight, age splits, etc.).
- label*: A descriptive label of the covariate values that were used to compute the effects (e.g., 5 years, 12 kg, below median weight [40–75](kg), ...).
- mid*: The point estimate.
- lower*: The lower end of the computed interval.
- upper*: The upper end of the computed interval.
- LABEL*: The formatted text string that is presented in the table (not needed when using the R Shiny application as it is generated automatically from the mid, lower, and upper).

Full documentation about the function use and examples can be found by running the command: “*help(forest_plot)*.” The included supplementary materials walk the user through the step-by-step process of the simulations and construction of the datasets used as input for the *cov-effectsplot forest_plot* function.

To ease the generation of the plot, all design aspects of the forest plot that can be configured interactively by invoking the R Shiny application included in the package by running one of these commands:

1. “*run_interactiveforestplot(yourdataname)*”: Launch the app with the dataset already available in the R session.
2. “*run_interactiveforestplot()*”: Launch an R Shiny application without a dataset, the user can then navigate to a saved csv file to get to the data that will be used for the forest plot.

An example of the user interface after loading a built-in example dataset is presented in Figure 7.

The R Shiny application allows the user to easily control the inclusion or exclusion of parameters or covariates, to control the order of presentation of parameters, covariates, and covariate value labels using “drag and drop” interactions. The application provides flexibility on formatting, with the possibility of changing font sizes, colors, ordering of legends, positioning of the facet strips, x and y axes titles, and more. The main graphical interface options and features are presented in Supplementary Material S2 to S6. After selecting the style of presentation using point-and-click interactions, the user can get the underlying R code by clicking the “show code” button below the forest plot. This helps the user in the learning process of a complex function with many arguments and ensures that the results can be reproduced by rerunning

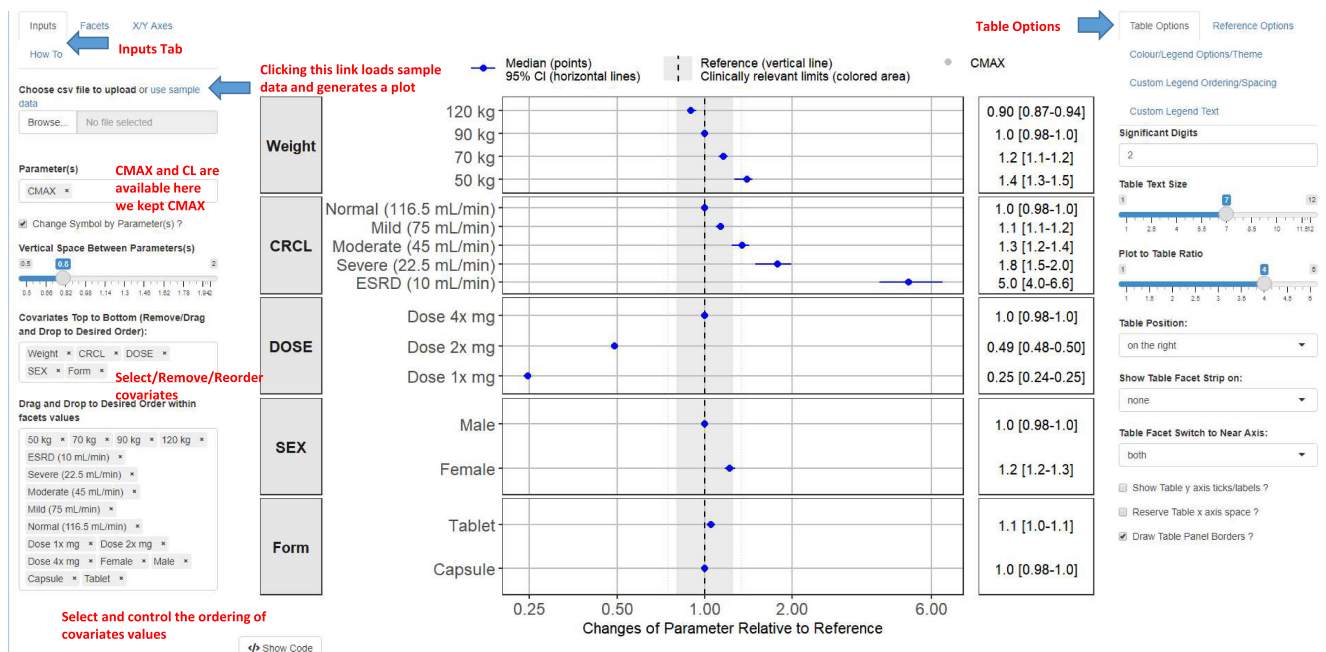


FIGURE 7 The R Shiny app user interface: inputs and table options tabs. CI, confidence interval; C_{max} , maximum plasma concentration; CRCL, creatinine clearance; ESRD, end-stage renal disease.

the code. Supplementary Material S7 show a screenshot of how this functionality works. Advanced R users, also have the capability to further fine tune the underlying “*ggplot2*” objects by asking the function to return a list of objects: `forest_plot(...,return_list = TRUE)`.

The application includes a “how to” menu, which provides detailed information regarding data structure and links to the package vignettes that guide the user into step-by-step tutorials and examples. The *coveffectsplot* package version 1.0.2 was used for the generation of the forest plots examples presented in the current tutorial. The package source code is hosted on Github <https://github.com/smouksassi/coveffectsplot>. The package also includes additional example use cases: PK/PD and exposure response not presented in this tutorial.

Throughout the examples, we did not discuss the R code related to the simulations, per se, data manipulation and summarization parts as these steps are not specific to the *coveffectsplot* package. Several excellent and well-documented R-based simulation tools with a lot of materials and tutorials that can be easily found online. The simulation engine used for all examples is all examples is the R package *mrgsolve*.¹³ The user has the freedom to bring in computed intervals data from any software or R code workflow available to him.

DISCUSSION

There is currently a need to develop software framework and code workflows for assessing and interpreting covariate effects in pharmacometrics projects, as this is required in all pharmacometric projects submitted to regulatory agencies. Although the discussed techniques are common, and are already part of the pharmacometricians’ toolkits, there is currently a lack of software tools that can help to implement and automate all the required steps. The focus of the *coveffectsplot* package and associate R Shiny application is to remove technical hurdles that can prevent the generation of covariate effects intervals and their presentation as forest plots for an effective communication of covariate effects.

Consistent with the US Food and Drug Administration (FDA) guidance for industry on population PK analyses,⁶ examples of simulations incorporating uncertainty on fixed-effect parameters only, or with BSV and uncertainty are presented in the current tutorial. The workflow based on uncertainty of fixed-effect parameters only is less computer intensive and provides a quick overview on the relative importance of covariates effects on exposure parameters. The drawback of this approach is that non-plausible scenarios can be obtained by varying covariates one at a time. Although this approach can help in isolating

the effects and relative importance of each covariate, a possible enhancement is to have values of interest conditional on a subpopulation of interest (e.g., body weight cutoff within male and female patients and within age groups).

Simulations based on uncertainty of fixed-effect parameters, BSV, and uncertainty on BSV is considered more robust and realistic, as it provides the joint effects of BSV and multiple covariates based on a database of real patients or in a virtual population with correlated covariates. The drawback of this approach is that it is computer and memory intensive and depends on the availability of a large distribution of covariates in the population. As a result, adequate planning of resources, optimization of code, and using modern, fast, and memory efficient tools, like the *mrgsolve*¹³ and *data.table*¹⁴ are essential when implementing a simulations framework based on uncertainty of fixed-effect parameters, BSV, and uncertainty on BSV.

In this tutorial, we did not simulate with the residual error (and uncertainty on residual error) because the goal was not to understand the range of observed concentrations in future populations. The proposed software framework can be easily extended by enabling the residual error with or without uncertainty.

To be able to understand and interpret intervals presented in forest plots, all the details describing the methodology and underlying simulations framework should be provided and the shown intervals need to be viewed in the context of the drug efficacy and safety.

In conclusion, the proposed simulation workflows have value and the *coveffectsplot* R package is expected to standardize and streamline the assessment and the communication of the covariate effect in pharmacometric projects. The package’s purpose is to facilitate clear data presentation that can support an obvious message. This in turn can accelerate the execution of projects due to better communication within drug development teams and efficient communication with regulatory agencies as evidenced by multiple submissions and approvals that have used this approach.¹⁵

ACKNOWLEDGEMENTS

The authors would like to acknowledge the contribution of Dean Attali for the development of the R Shiny Application and Benjamin Rich for insightful comments on the R package functions and for writing the function *expand.modelframe*. We also thank the editor for inviting us to submit this tutorial, and two anonymous reviewers for very informative comments that significantly improved the manuscript over the review cycles.

CONFLICT OF INTEREST

The authors declared no competing interests for this work.

ORCID

Jean-Francois Marier  <https://orcid.org/0000-0001-9820-5116>

Nathan Teuscher  <https://orcid.org/0000-0002-4133-6374>

Mohamad-Samer Mouksassi  <https://orcid.org/0000-0002-7152-6654>

REFERENCES

- Mehrotra S, Gobburu J. Communicating to influence drug development and regulatory decisions: a tutorial. *CPT Pharmacomet Syst Pharmacol*. 2016;5:163-172.
- Vandemeulebroecke M, Baillie M, Margolskee A, Magnusson B. Effective visual communication for the quantitative scientist. *CPT Pharmacomet Syst Pharmacol*. 2019;8:705-719.
- Mouksassi S, Attali D. `coveffectsplot`: Produce Forest Plots to Visualize Covariate Effects. <https://cran.rstudio.com/web/packages/coveffectsplot> (2022). Accessed 1 April 2022.
- Freiman JA, Chalmers TC, Smith H Jr, Kuebler RR. The importance of beta, the type II error and sample size in the design and interpretation of the randomized control trial. Survey of 71 "negative" trials. *N Engl J Med*. 1978;299(13):690-694.
- Lewis JA, Ellis SH. A statistical appraisal of post-infarction beta-blocker trials. *Prim Cardiol*. 1982;suppl 1:31-37.
- Guidance for Industry. Population Pharmacokinetics. February 2022. <https://www.fda.gov/media/128793/download>. Accessed 1 April 2022.
- Menon-Andersen D, Yu B, Madabushi R, et al. Essential pharmacokinetic information for drug dosage decisions: a concise visual presentation in the drug label. *Clin Pharmacol Ther*. 2011;90:471-474.
- Xu XS, Yuan M, Zhu H, et al. Full covariate modelling approach in population pharmacokinetics: understanding the underlying hypothesis tests and implications of multiplicity. *Br J Clin Pharmacol*. 2018;84:1525-1534.
- Leeper TJ. Interpreting regression results using average marginal effects with R's margins. <https://cran.r-project.org/web/packages/margins/vignettes/TechnicalDetails.pdf> (2018). Accessed 12 December 2021.
- Ludecke D. Plotting marginal effects of regression models. https://cran.r-project.org/web/packages/sjPlot/vignettes/plot_marginal_effects.html (2021). Accessed 1 April 2022.
- Center for Disease Control and Prevention (CDC). <https://www.cdc.gov/growthcharts/data/zscore/wtageinf.csv> (2009). Accessed 1 April 2022.
- Stasinopoulos M, Rugby R, Akantziliotou C, et al. `gamlss.dist`: distributions for generalized additive models for location scale and shape. <https://cran.r-project.org/web/packages/gamlss.dist/> (2022). Accessed 1 April 2022.
- Baron KT, Gillespie B, Margossian C, et al. `mrgsolve`: simulate from ODE-based models. <https://CRAN.R-project.org/package=mrgsolve> (2022). Accessed 1 April 2022.
- Dowle M, Srinivasan A, Gorecki J, et al. `Data.table`: extension of 'data.frame'. <https://cran.r-project.org/web/packages/data.table/> (2021). Accessed 1 April 2022.
- Drug approval package: `onpattro` (patisiran): multi-discipline review/summary, clinical, non-clinical. https://www.accessdata.fda.gov/drugsatfda_docs/nda/2018/210922Orig1s000MultiR.pdf. Accessed 22 March 2022.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Marier J-F, Teuscher N, Mouksassi M-S. Evaluation of covariate effects using forest plots and introduction to the *coveffectsplot* R package. *CPT Pharmacometrics Syst Pharmacol*. 2022;11:1283-1293. doi:[10.1002/psp4.12829](https://doi.org/10.1002/psp4.12829)