Virtual patient simulation using copula modeling

Zwep, L.B.; Guo, T.; Nagler, T.W.; Knibbe, C.A.J.; Meulman, J.J.; Hasselt J.G.C. van

Virtual Patient Simulation Using Copula Modeling

Laura B. Zwep[1], Tingjie Guo[1], Thomas Nagler[2], Catherijne A.J. Knibbe[1,3], Jacqueline J. Meulman[4,5] and J. G. Coen van Hasselt[1,*]

Virtual patient simulation is increasingly performed to support model-based optimization of clinical trial designs or individualized dosing strategies. Quantitative pharmacological models typically incorporate individual-level patient characteristics, or covariates, which enable the generation of virtual patient cohorts. The individual-level patient characteristics, or covariates, used as input for such simulations should accurately reflect the values seen in real patient populations. Current methods often make unrealistic assumptions about the correlation between patient's covariates or require direct access to actual data sets with individual-level patient data, which may often be limited by data sharing limitations. We propose and evaluate the use of copulas to address current shortcomings in simulation of patient-associated covariates for virtual patient simulations for model-based dose and trial optimization in clinical pharmacology. Copulas are multivariate distribution functions that can capture joint distributions, including the correlation, of covariate sets. We compare the performance of copulas to alternative simulation strategies, and we demonstrate their utility in several case studies. Our work demonstrates that copulas can reproduce realistic patient characteristics, both in terms of individual covariates and the dependence structure between different covariates, outperforming alternative methods, in particular when aiming to reproduce high-dimensional covariate sets. In conclusion, copulas represent a versatile and generalizable approach for virtual patient simulation which preserve relationships between covariates, and offer an open science strategy to facilitate re-use of patient data sets.

Study Highlights

WHAT IS THE CURRENT KNOWLEDGE ON THE TOPIC?

☑ Trial and dose optimization for different types of patient populations can be informed simulations from pharmacokinetic and pharmacodynamic models. Typically, this virtual patient simulation requires simulation of realistic patient characteristics and combinations of these characteristics.

WHAT QUESTION DID THIS STUDY ADDRESS?

☑ How can copulas be used for simulation and sharing of realistic patient data?

WHAT DOES THIS STUDY ADD TO OUR KNOWLEDGE?

☑ Copulas are able to simulate realistic patient characteristics in higher dimensions. The distribution-based approach allows for the interpolation to new patient populations of interest, and sharing of simulations and creating simulation tools.

HOW MIGHT THIS CHANGE CLINICAL PHARMACOLOGY OR TRANSLATIONAL SCIENCE?

☑ Copula simulations can assist in clinical trial development and dose optimization through *in silico* studies of different virtual (special) patient populations, by retaining the dependence structure between patient's covariates. The copula approach allows for sharing of simulated populations and simulation tools, removing the need to share the underlying patient data.

Model-based approaches in pharmacometrics and quantitative systems pharmacology (QSP)[1–3] have become of pivotal importance for the optimization of drug treatment strategies or clinical trial designs.[4,5] These model-based approaches typically simulate the expected pharmacokinetic (PK) and/or pharmacodynamic (PD) response and the associated interindividual variability for a cohort of virtual patients. Here, the interindividual variability in the PK or PD response is often in part captured by patient-specific characteristics, such as age, weight, organ function biomarkers, or specific genetic polymorphisms, incorporated in quantitative PK-PD or QSP models. The increasing public availability of quantitative PK-PD or QSP models for many important therapeutics thus

[1]Division of Systems Pharmacology and Pharmacy, Leiden Academic Centre for Drug Research, Leiden University, Leiden, The Netherlands; [2]Department of Statistics, Ludwig Maximilian University of Munich, Munich, Germany; [3]Department of Clinical Pharmacy, St. Antonius Hospital, Nieuwegein, The Netherlands; [4]LUXs Data Science, Leiden, The Netherlands; [5]Department of Statistics, Stanford University, Stanford, California, USA. *Correspondence: J.G. Coen van Hasselt (coen.vanhasselt@lacdr.leidenuniv.nl)

Received June 27, 2023; accepted October 31, 2023. doi:10.1002/cpt.3099

offers extensive opportunities for the clinical pharmacology community to perform virtual patient simulations. These simulations may aid in design of (stratified) dosing strategies, in particular for new (special) patient population populations,[6] such as pediatric[7–9] or pregnant patients,[10] or to evaluate different potential trial designs in specific types of patients or treatments[11–13] (**Figure 1a**).

A key requirement to enable simulation of realistic virtual patients is to produce realistic sets of patient-associated characteristics or covariates used in the model. Such covariates can include demographics (e.g., body weight, and sex, age), organ function measures (e.g., renal or hepatic function), PD end points (cardiovascular readouts and biochemical biomarkers), and increasingly also high-dimensional pharmacogenomic data. Importantly, such covariates may have various distributions, including an intricate dependency structure (i.e., correlation) that must be accounted for in virtual patient simulation to produce realistic patient-profiles (**Figure 1b**). Not considering such correlations leads to an inflation of the variability in covariates and hence unrealistic virtual patients. For example, a patient of 95 years old, with a high body weight and a very good kidney function is a combination that is not expected to actually exist. Next to more mechanistic simulation methods,[14] various data analytical strategies are available to generate sets of realistic patient covariates for virtual patient simulation. These strategies are either based on methods that require direct access to the appropriate individual patient-level covariate data, which may often not be available, or on methods that characterize the covariate distributions.

Covariate generation methods that utilize available patient-level covariate data include resampling methods, such as the bootstrap,[15] which preserve the dependence structure of the patient covariates by directly resampling from the observed data. However, these methods are only able to simulate patients that are already present in the data set and require a large enough number of patients to be included. These shortcomings were addressed by a recently proposed imputation method using conditional distributions (CDs),[16] although this method remains dependent on access to patient-level data. Distribution-based simulation methods for virtual patient simulation do not require patient-level data access. Although initially distributions are often derived from patient-level data, subsequent use of these distributional models to generate sets of patient-level covariates is independent of access to such data. The most straightforward strategy is to capture the marginal density of covariates in univariate parametric distributions with associated means and variances for each covariate, and to subsequently draw random samples from these distributions. However, such an approach assumes that covariates are fully independent and do not show any correlation. Alternatively, multivariate normal distributions (MVNDs)[17] do capture the correlation structure,[18] but make strong assumptions regarding the (multivariate normal) distributional shape, which is commonly violated. Thus, depending on the distribution of the covariates of interest this again can lead to unrealistic sets of virtual patient covariates.

Copulas are multivariate distribution functions that can capture the joint distribution, including the dependence structure for sets of covariates, and are thus of interest as a distribution-based approach for generating realistic sets of covariates. They address

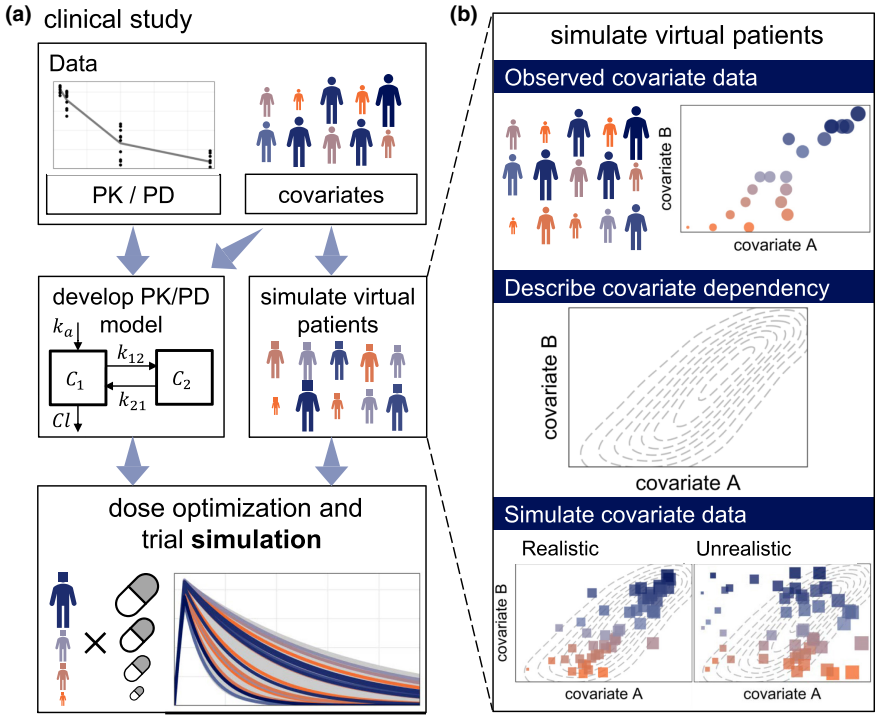

Figure 1 Pharmacometric workflow. (**a**) In order to optimize dosing for new medication or special patient populations, pharmacometric models, such as PK/PD models, are used to simulate new patient dosing regimens. Next to the developed pharmacometric model, simulation studies require covariate simulation. (**b**) An important challenge for covariate simulation is sampling realistic patients, where the dependency between covariates is preserved. PD, pharmacodynamic; PK, pharmacokinetic.

shortcomings of alternative distribution-based methods while not requiring access to patient-level data.[19–21] In this study, we aim to evaluate and demonstrate the utility of copulas as a novel strategy to support realistic virtual patient simulation in the context of the field of clinical pharmacology. We first compare the performance of copula models in comparison to existing methods, including the bootstrap, CD, MVND, and marginal distribution. We then demonstrate the application of copulas in three case studies focusing on PK simulations, time-varying covariates, and higher-dimensional covariates.

METHODS

Data

Three different data sets of combined patient characteristics were used in this study to evaluate the performance and explore different applications (**Figure S1**). The first data set contains a special patient population of pediatric patients[22] with 445 neonates and young children admitted to the intensive care unit (ICU), with 12 measured covariates, including body weight, serum creatinine (SCr) level, and age. These data were used to evaluate the simulation performance (data set 1, **Table S1**). A second data set on pregnancy data[23] with 123 subjects, with biomarkers measured over time, was used to simulate longitudinal covariate profiles (data set 2, **Table S2**). Last, MIMIC,[24] a large observational data set with ICU patients, was used to evaluate the correlation structure between a large set of 30 variables for > 53,000 patients (data set 3, **Table S3**).

Copula estimation and simulation

Vine copulas were used to estimate the joint density between all covariates. Vine copulas are multivariate densities constructed through a combination of copula pairs, which increases flexibility and reduces the computational burden for the estimation of the copulas.[25] Kernel density estimation was used to estimate the marginal density of each covariate. Using the probability integral function, the covariates were transformed to a uniform scale, with values on the [0,1] domain.[26] Based on the correlations between the covariates, a vine structure was chosen, where the most correlated covariates were placed closer to each other in the vine structure.[27] For each bivariate copula, a set of parametric distributions was fit and the best fitting distributions were chosen by minimizing the Akaike information criterion (AIC). Vine copulas with different distributions were fit using the R library *rvinecopulib*.[28] The resulting copula density was used to simulate covariates with uniform marginal densities. The earlier estimated marginal densities were used to transform these covariates back to their original scale, yielding the simulated covariate sets for virtual patients. All analyses were performed in R. Scripts and estimated copulas are available on GitHub (https://github.com/vanhasseltlab/copula_vps).

Evaluation of simulation performance

To evaluate how well copulas can be used for simulation of covariate sets, we calculated the performance of copula simulations on the pediatric data[22] (data set 1). The estimation and simulation were performed in 2 differently sized covariate sets, with the same subjects, but a different number of covariates: one simulation on 3 covariates, age, SCr, and body weight, and one on 12 covariates. The distribution of the simulated population was compared with the distribution of the observed population in terms of the sample mean and sample standard deviation for each covariate and the sample correlation between each combination of covariates. A relative error was computed for each of these statistics (S) as

$$\text{Relative error} = \frac{\hat{S} - S}{S},$$

where $\hat{S}$ denotes the statistic of the simulated population. The simulations were repeated 100 times.

The copula results were compared with four other simulation methods, of which two methods are based on patient-level data and two methods are based on characterization of the covariate distribution. Bootstrap simulations were conducted by resampling full rows from the original data with replacement.[15] The CD approach, which uses a multiple imputation algorithm to iteratively impute covariate values for virtual patients, was used as implemented by the developers of the method.[16] The standard multiple imputation method "predictive mean matching" was used, corresponding to their paper. The distribution-based methods used were the MVND and marginal distributions (MDs), through maximum likelihood estimation. The best fitting multivariate normal distribution was fitted. The univariate MDs of each covariate was estimated using a kernel density estimation method.[26,29] Covariate values were sampled from the respective density functions.

Applications

Pharmacokinetic simulation of vancomycin in pediatric patients. For the proposed copula approach, the effect of preserving the dependence structure in covariate simulation methods was evaluated on PK predictions in pediatric patients. To this end, for data set 1, the performances of the use of body weight and SCr from the three-covariate copula, the MDs, and the MVND simulation were compared in a population PK one-compartmental model for vancomycin.[30] In case of the MVND, the covariates were log-transformed before fitting an MVND, to ensure non-negativity while simulating the body weight and SCr values. The simulated values were back-transformed to original scale.

$$\frac{dA}{dt} = k_o - \frac{CL}{V} \cdot A$$

$$CL = \frac{3.56 \cdot WT}{SCr}$$

$$V = 0.669 \cdot WT$$

This PK model was used to calculate the PK curves from the original pediatric covariate data (data set 1) and the simulated covariate data from the three-covariate copula and MD simulations. These PK profiles were compared using the area under the curve (AUC) of the first 24 hours after dosing, calculated using a trapezoidal method. The correlation between the AUC and the covariates, and the SCr and body weight, was evaluated to identify whether this correlation was recovered between the covariates and the PK curve.

Time-varying covariates in pregnancy data. One of the possible applications of using copulas is the simulation of time-varying covariates. Using data set 2 with 6 time-varying covariates (y) over the gestational age (t) during pregnancy,[23] including albumin concentration, bilirubin concentration, lymphocytes, neutrophils, platelets, and SCr, we fit a copula to simulate time-varying covariates in a two-step procedure. First, we fitted a second degree mixed effects polynomial regression model on the temporal data for each covariate j and extracted 3 individual parameters for each patient i, the intercept $(\beta_{0j} + b_{0ji})$, the linear term $(\beta_{1j} + b_{1ji})$, and the quadratic term $(\beta_{2j} + b_{1ji})$, resulting in a total of 18 dimensions.

$$\hat{y}_{ij}(t) = \beta_{0j} + b_{0ji} + \beta_{1j} \cdot t + b_{1ji} \cdot t + \beta_{2j} \cdot t^2 + b_{2ji} \cdot t^2$$

$$b_{0ji} \sim N(0, \sigma_0)$$

$$b_{1ji} \sim N\left(0, \sigma_1\right)$$

$$b_{2ji} \sim N\left(0, \sigma_2\right)$$

For example, yielding for albumin concentration:

$$\widehat{\text{Albumin conc}}_i(t) = 44.1 + b_{0i} \pm 0.269 \cdot t + b_{1i} \cdot t + 0.0017 \cdot t^2 + b_{2i} \cdot t^2$$

$$b_{0i} \sim N(0, 1.86)$$

$$b_{1i} \sim N(0, 0.105)$$

$$b_{2i} \sim N(0, 0.00224)$$

Second, instead of fitting a copula directly on the longitudinal covariates, the copula was fitted on the set of individual parameter estimates, yielding the six new sets of intercepts, linear, and quadratic terms for each simulated patient. To create time-dependent covariates, the curves for each patient were retrieved from the simulated parameter sets. The performance of the copula simulation was evaluated by comparing the time-curves estimated from the copula simulated time curves with those estimated on the original pregnancy data. The performance was evaluated both in terms of the simulated individual parameters as the calculated time-curves. Next to simulation with the copula, the time-varying covariates were simulated in a similar two-step approach with MDs, to compare the differences between the MDs and copula.

Covariate distributions in large ICU data. To characterize the joint distributions in a large data set, copula simulation was used to characterize and simulate from the MIMIC database (data set 3).[24] A copula model was fit to a large data set of 30 available patient-associated covariates with primary focus on clinical laboratory measurements from > 53,000 ICU patients. There were many values missing over the covariates and subjects. To estimate the copula on missing data, for each combination of covariates needed for a node in the vine copula structure, the complete observations were used. This simulation was used to demonstrate how copulas can be used to characterize the underlying dependency structure of these covariates and evaluate the correlations.

RESULTS

Evaluation of simulation performance

The performance of the copulas was assessed on 2 differently sized data sets, one with 3 covariates, and one with 12 covariates (data set 1). First, for a set of 3 covariates, copulas show a low relative error of −0.02, −0.08, and −0.04 for the terms of correlations between age and body weight, age and SCr, and body weight and SCr, respectively (**Figure 2a**). Second, for the 12-covariate simulations, the copula simulation slightly underestimates the covariances with a median error of −0.05 over all covariate combinations (**Figure 2b**).

The performance of copulas was compared with four other simulation methods. For the 3-covariate simulation the copula yielded similar results to the conditional distributions, which has relative errors of −0.01, −0.12, and −0.03 (**Figure 2a**), but for the 12-covariate simulations, the CD simulations show a large median underestimation with a relative error of −0.60 (**Figure 2b, Table S4**). The bootstrap shows the best performance, because it can fully keep the dependence structure intact, both in the 3-covariate (**Figure 2a**) and the 12-covariate simulation (**Figure 2b**). The MDs was unable to capture any correlation, which is seen in the relative error of around −1.0 for each covariate combination. The MVND shows a good performance in the estimates for correlation, mean, and standard deviation, but a visual check of the density plots shows a non-normal distribution of the covariates, which is not well covered by the simulated density (**Figure S2**).

Overall, copulas performed closest to the bootstrap, which can fully capture the dependence, but it was not able to capture all covariate combinations equally well, such as a large overestimation of the combination CREF and FRCR. The 12-covariate model showed a weakness in the conditional distributions, which the copulas did not show and, although the MVND shows very good summary metrics, the distributions themselves perform worse than the copula (**Figure S2**).

Applications

Pharmacokinetic simulation of vancomycin in pediatric patients. The effect of ignoring the correlation between covariates on PK simulations was evaluated by comparing the PK curves from the copula simulations with those from the MDs simulation. Covariate sets simulated for SCr and body weight from data set 1 were used to predict PK profiles and compute subsequent AUCs. The AUCs from the copula and the MDs simulations did not show differences in summary statistics, such as the median and quartiles (**Figure 3a**). For the log-MVND, a part of the simulated values was outside of the range of the original data and the normal distribution lead to a lower median AUC and higher 97.5 percentile. However, when comparing the correlations between the covariates and the AUC, we found that the original correlation between the AUC and body weight ($r = -0.67$) was lost in the MD simulations ($r = -0.07$), whereas the copula ($r = -0.66$) and the MVND ($r = -0.58$) mostly preserved their dependence (**Figure 3b**). If the dependence between variables is not taken into account, this can lead to unrealistic virtual patients, such as individuals with a high body weight having a high AUC.

Time-varying covariates. To evaluate how well copulas can be used to simulate time-varying covariates, a two-step simulation method was used to simulate patients, with and without taking the dependency into account, by simulating from a copula and MDs, respectively. For the time-varying covariates in the pregnancy data (data set 2), polynomial linear regression curves were fitted for each covariate, resulting in polynomial equations. The individual parameters were estimated, resulting in a set of 18 parameter estimations for all subjects. A set of virtual patients was simulated from the estimated individual parameters. The correlations between the individual parameters from the simulated patients were on average close to the correlations between the estimated parameters of the observed data. The simulated individual parameters were used to generate time-varying covariate values, by calculating the curves from the intercept and the linear and quadratic terms. Polynomial regression coefficients were simulated in a realistic

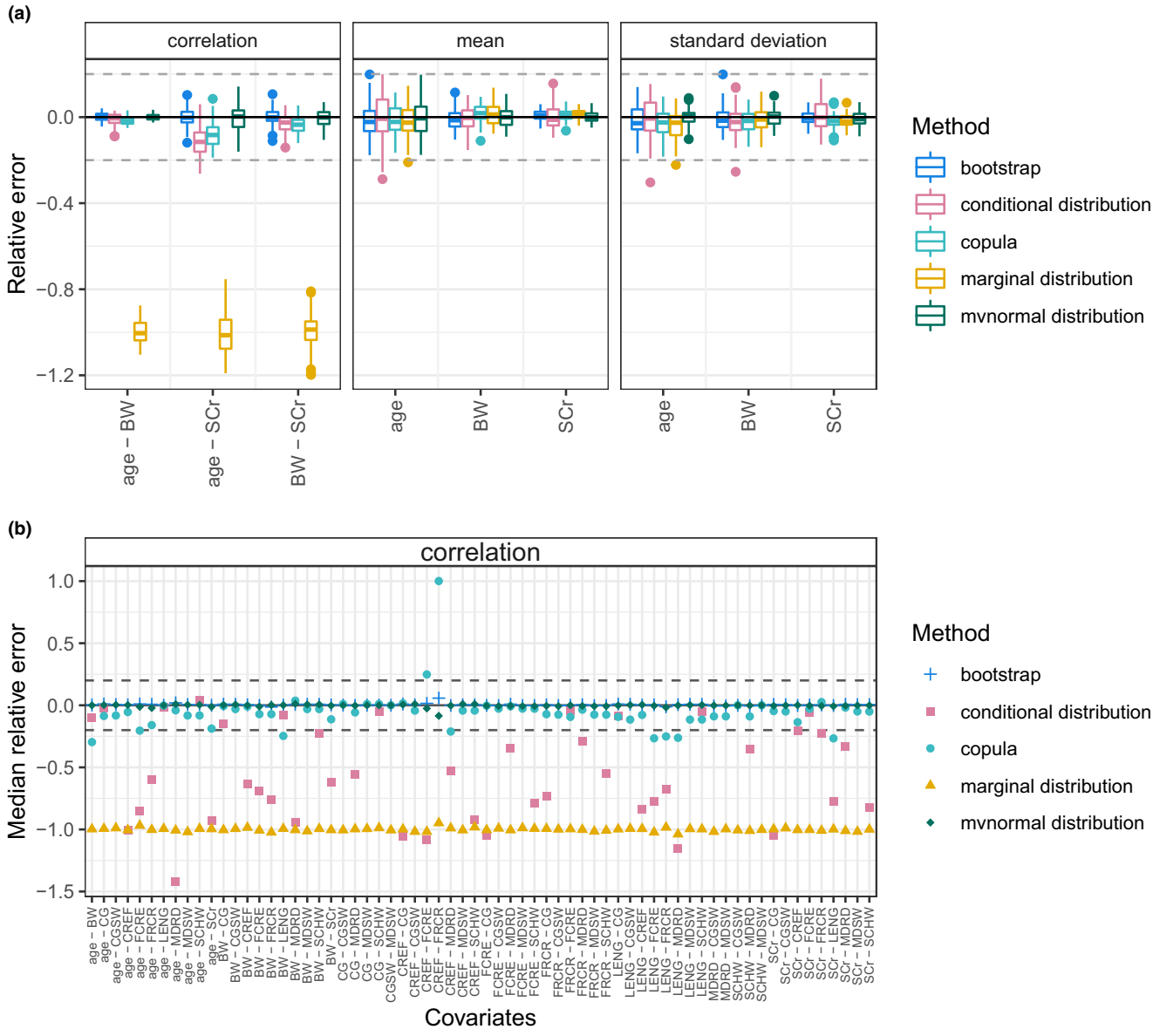

Figure 2 Relative error over 100 simulations as compared with the statistics of the observed population for five different simulation methods. (**a**) Boxplots of the correlation, mean and standard deviation of three covariates. (**b**) Median relative error of a large covariate simulation for the correlations of each combination of 12 covariates. BW, body weight; SCr, serum creatinine.

range, apparent from the calculated polynomial curves which overlap the observed polynomial curves, whereas simulating from an MD led to more extreme polynomial curves, with a five times higher error on the standard deviation of the AUC (**Figure 4**). This shows how covariate values can be inflated when simulating independent covariates.

Covariate distributions in large ICU data. To establish the use of copula for simulation in a larger data set, a simulation was conducted based on 30 covariates from the MIMIC database (data set 3). Copula estimation and simulation was feasible on this large data set, showing how copulas can be useful for simulation for extensive pharmacometric models. The higher dimension did increase the underestimation of the correlations to a relative error of −0.11, which was slightly worse compared with the estimation in the lower dimensional 12- and 3-covariate data sets (**Figure S3**). Some covariates show interesting dependency structures, which can be evaluated and be used in covariate selection decision making (**Figure 5**). The results from the larger data set also show that through the use of copulas, it is feasible to share hospital data distributions.

DISCUSSION

We showed a competitive or superior performance of copula simulations compared with other simulation methods, and we demonstrated multiple applications for covariate simulations using copulas. Copulas were able to preserve the correlations between covariates in lower and higher dimensional data sets. Preserving the dependence structure in copula simulations allows for simulating covariate sets for realistic PK predictions, time-varying

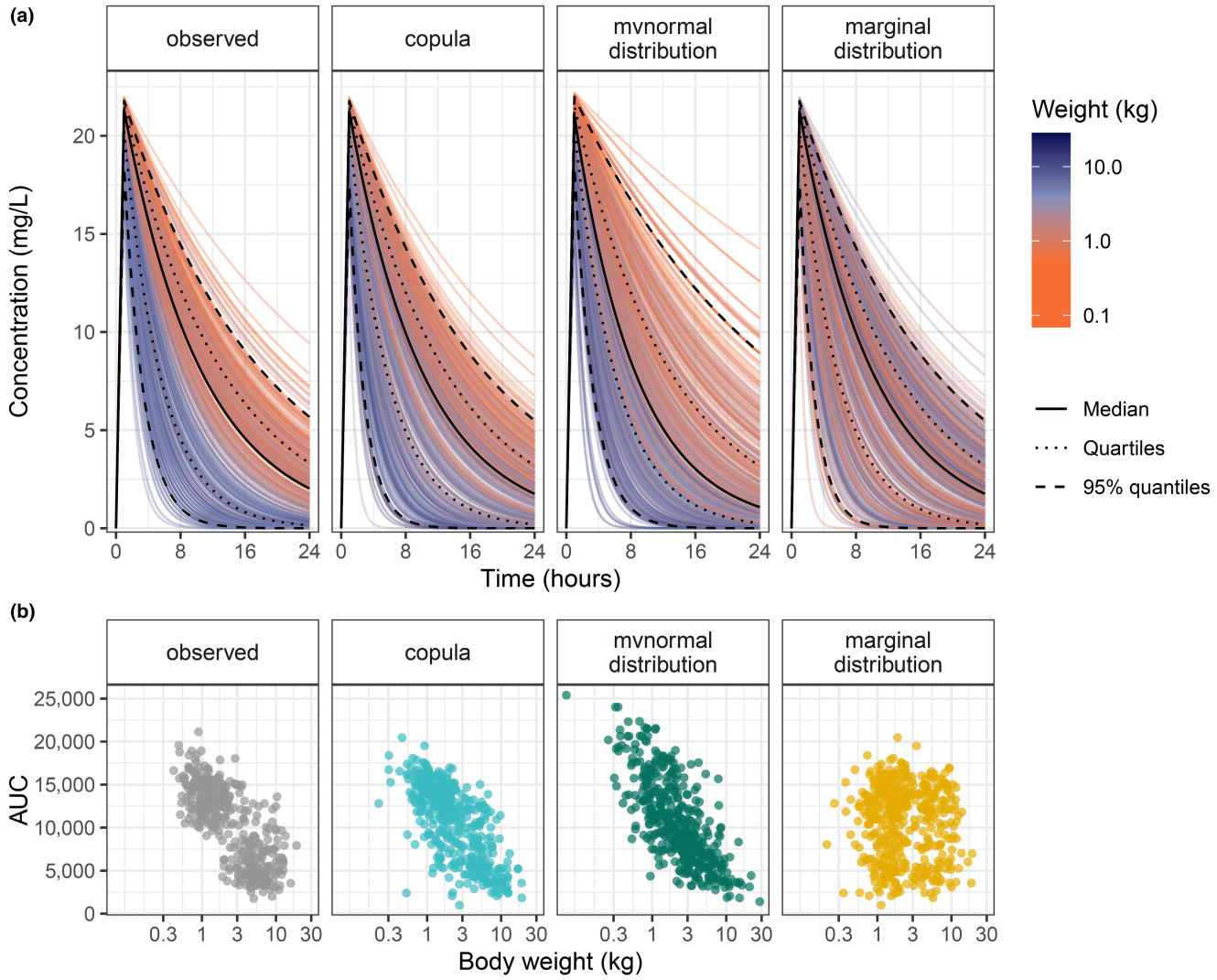

Figure 3 (**a**) Pharmacokinetic (PK) curves calculated for the observed population and the virtual patient populations from copula simulations, the multivariate normal distribution (MVND) and the marginal densities (MDs). The median and quantiles show a similar pattern between copula and MDs, however, the weight is randomly distributed over the PK profiles for the simulation with MDs. The MVND shows the same correlation, but a slightly different pattern in the median and quantiles. (**b**) Scatter plot of area under the PK curve (AUC) against body weight (log-scale).

covariates, and in a large-scale data set, that is, the MIMIC data, thus making it a suitable method for virtual patient covariate simulations in a variety of settings. Copula simulation has apparent benefits over currently used methods, because these either neglect the dependence structure among the covariates, the shape of their distribution, or rely on real patient data in simulation.

We evaluated the performance of copulas compared with other simulation methods. Although performing well in lower dimensions, we observed increasing underestimation in higher dimensions for CD, making the method less suitable for simulations in higher dimension, which is an increasingly important feature, due to the rise in models which include multiple biomarkers and clinical covariates.[31] The MVND showed very promising results in terms of capturing the correlation (**Figure 2**). However, this is an inherent feature of how the MVND is estimated, which is based on the mean, standard deviation, and covariance. It does, on the other hand, not capture the actual shape of the distribution when covariates are not normally distributed (**Figure S2**). Although the bootstrap can fully preserve the dependence structure between covariates, it cannot be used for simulation when actual data are unavailable. Additionally, due to the resampling nature of the bootstrap, one cannot simulate covariate values for virtual patients beyond which are present in the actual data set, which may result in simulating an unbalanced virtual patient population. Although copulas require the use of underlying data to be estimated, the simulated covariate values and the joint density functions can be shared without including any patient information, making it possible to publish the simulated data sets and simulator. The application of MDs was shown to simulate unrealistic patients, in the three situations studied.

Preserving the dependence between covariates is required for simulation of realistic patients in terms of PK predictions in the pediatrics vancomycin model, used in this study. The copula was able to preserve the relationship between the body weight and the

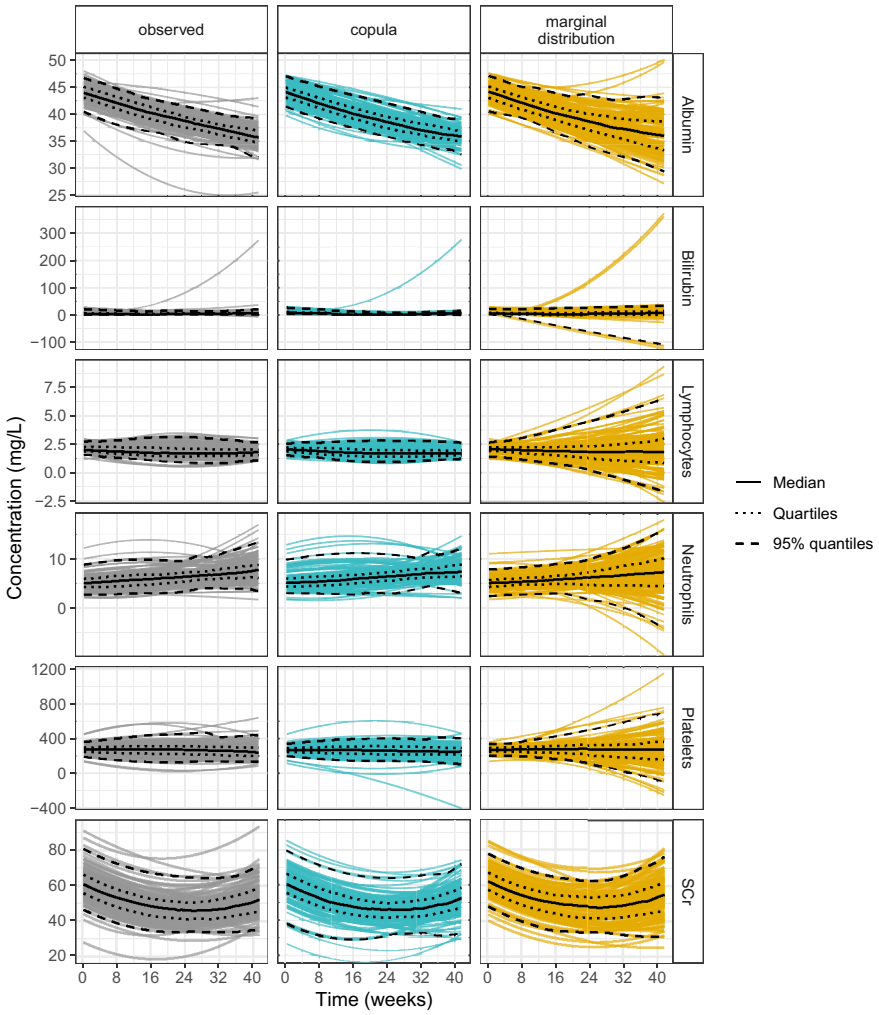

Figure 4 Polynomial curves for the six biomarkers from pregnancy data. In gray are the estimated curves from the observed data. The copula (turquoise) shows very similar patterns, whereas the marginal distribution (yellow) shows extreme values, especially at the end of the curve. SCr, serum creatinine.

AUC, which is of high clinical relevance. This feature of copulas provides a significant insight into how PKs may differ between subgroups of patients. It allows one to optimize the dose for a particular patient group or to study the differences between patient groups. We found that PKs at the population level is not affected by the method used for virtual patient simulation (**Figure 3**). This is expected due to the negative correlation between SCr and body weight and the inversely proportional relation between these covariates in their effect on the PKs. An unrealistic individual, with high body weight and high SCr, would still have a PK profile within a realistic range. However, the impact of preserving the dependence structure can differ per model, as can be seen in simulating the time-dependent covariates in the analysis of the pregnancy data. Here, polynomial regression coefficients need to be simulated in a realistic domain, in order to preserve the structure of the data, both on the individual and population levels. Simulating from a marginal distributions lead to extreme polynomial curves.

Access to real individual-level patient data is often hampered by personal data protection regulations, which is a significant obstacle for community-driven design of optimized treatment strategies and trial designs.[32] Although copulas are mostly estimated on data, resulting copulas can be easily shared without sharing patient data, allowing one to use established copulas for virtual patient simulation.[33] Using copulas both opens opportunities for better replication and comparison studies, and copulas can facilitate in simulation platforms for sharing patient characteristics. As such, it is possible to simulate patients from the data used in this study using the vine copula objects and scripts shared on GitHub.

In pharmacometric models, covariates typically only explain part of the interindividual variation, with the remaining variation accounted for by the random effects. Depending on the model, the contribution of the covariates in explaining variability can vary. Therefore, in order to appropriately simulate virtual patient cohorts using pharmacometric models, accurate modeling of interindividual random effects parameters and their correlation structure can be considered of equal importance.

The sharing of models has become more common in the pharmacometrics community, for example, through platforms for model sharing, such as DDMoRe. However, models often require covariate input. Copulas can be used to set up a large-scale

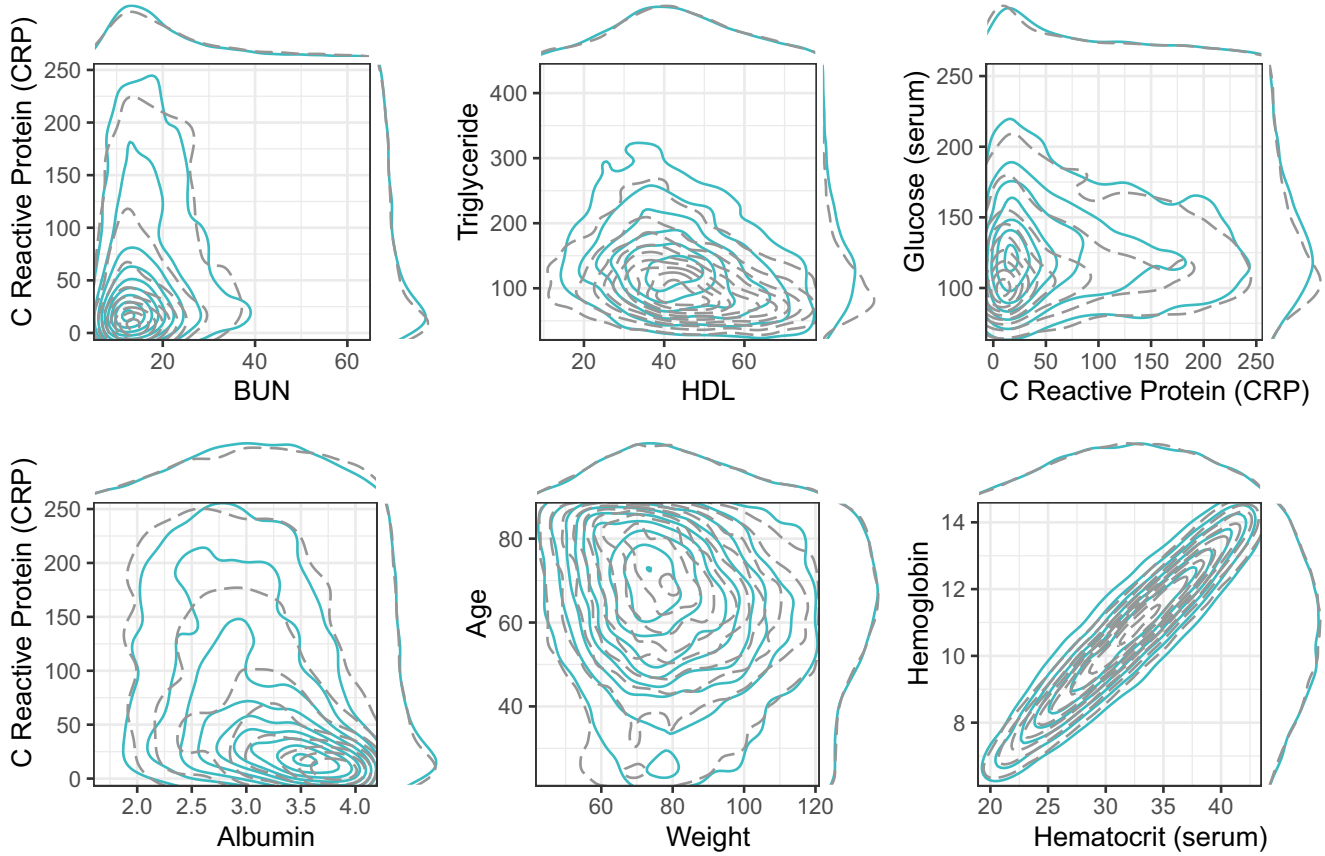

Figure 5 Set of selected covariate combinations with the densities of the observed population (gray dashed line) and the simulated population from a copula (blue solid lines), with marginal densities on the top and right sides of each plot. More overlap between the lines shows a better correspondence between the observed and simulated patient covariates.

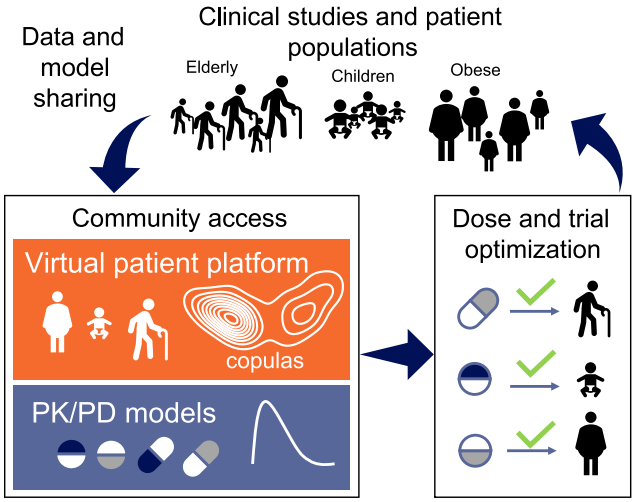

Figure 6 Community access pharmacometrics research pipeline. Data and pharmacometric models from (special) patient populations can be shared with the clinical pharmacology community. Through copulas, covariate sets can be simulated, which, when used in PK/PD models, can aid treatment and dosing optimization, ultimately improving treatment for the patients. PD, pharmacodynamic; PK, pharmacokinetic.

covariate simulation platform, which can accompany the shared models to allow the clinical pharmacology community to simulate clinical trials and dosing regimens for (special) populations, even when there is no patient-level data available (**Figure 6**). For these sharing opportunities, it is of interest to share a larger number of covariates, even though not all covariates are used in the same pharmacometric model.

The use of vine copulas allows for the estimation of flexible multidimensional densities. It, however, requires to choose a tree structure, which in this study was done using the AIC.[27] The AIC penalizes the size of the model, in terms of number of parameters in the distribution, which prevents always choosing the distribution with more parameters, which is the case when using the log-likelihood directly.[20]

This paper did not address simulation of categorical variables. Discrete, ordered categorical and binary covariates can be captured as a copula, by using rank-based distributions,[34] however, the copula method is not able to deal with unordered categorical variables in a natural way, because the used copula functions are monotonic.[35]

Regardless of the method of simulation, further research would also require looking into the underestimation of the correlation by the different simulation techniques, because there are limits to the full characterization of the joint distribution. Visualization of the simulation through density plots allows to investigate how severe the discrepancy between the observed and population and the copula is and whether it seems clinically relevant. This can be evaluated on the level of the covariates, but also by looking at the outcomes of pharmacometric models.[36]

In summary, copulas represent an attractive approach to capture multivariate covariate distributions, which can be readily implemented for pharmacometric simulations, including PK, PD, and QSP simulations. Copulas show superiority in the combination of being a flexible framework for adequately simulating covariates and being a tool useful for anonymous data sharing. The distribution-based nature of copulas has the distinct advantage that access to original individual-level data sets is not required when applied for virtual patient simulation, in contrast to resampling-based strategies. To this end, copula models can address hurdles in accessing real clinical data by developing open access simulation models for distinct (special) patient populations, which can be readily shared with the community and support clinical trial simulations and treatment optimization.

SUPPORTING INFORMATION

Supplementary information accompanies this paper on the *Clinical Pharmacology & Therapeutics* website (www.cpt-journal.com).

ACKNOWLEDGMENTS

The authors thank Matthijs de Hoog (Department of Pediatric Intensive Care, Erasmus MC - Sophia Children's Hospital, Rotterdam, The Netherlands), Karel Allegaert (Neonatal Intensive Care Unit, University Hospital Leuven, Leuven Belgium), Hussain Mulla (Department of Pharmacy, University Hospitals of Leicester, Leicester England, UK), and Catherine M. T. Sherwin (Department of Pediatrics, Wright State University Boonshoft School of Medicine/Dayton Children's Hospital. Dayton, OH, USA) for providing us with data on patient covariates in neonates/pregnant patients that was part of previous publications. We wish to also thank Jignesh P. Patel (Department of Hematological Medicine, King's College Hospital, London, UK) for providing us with previously published data on time-varying covariate data in pregnant patients.

FUNDING

No funding was received for this work.

CONFLICT OF INTEREST

The authors declared no competing interests for this work.

AUTHOR CONTRIBUTIONS

L.B.Z., T.G., T.N., C.A.J.K., J.J.M., and J.G.C.v.H. wrote the manuscript. J.G.C.v.H., L.Z., J.J.M., and T.G. designed the research. L.B.Z., T.G., and J.G.C.v.H. performed the research. L.B.Z. and T.G. analyzed the data.

1. Bonate, P.L. Clinical trial simulation in drug development. *Pharm. Res.* **17**, 252–256 (2000).
2. Bonate, P.L. A brief introduction to Monte Carlo simulation. *Clin. Pharmacokinet.* **40**, 15–22 (2001).
3. Chelliah, V. *et al*. Quantitative systems pharmacology approaches for immuno-oncology: adding virtual patients to the development paradigm. *Clin. Pharmacol. Ther.* **109**, 605–618 (2021).
4. Holford, N., Ma, S.C. & Ploeger, B.A. Clinical trial simulation: a review. *Clin. Pharmacol. Ther.* **88**, 166–182 (2010).
5. Langenhorst, J.B. *et al*. Clinical trial simulation to optimize trial design for fludarabine dosing strategies in allogeneic hematopoietic cell transplantation. *CPT Pharmacometrics Syst. Pharmacol.* **9**, 272–281 (2020).
6. Cock, P.A.J.G.D. *et al*. Population pharmacokinetics of cefazolin before, during and after cardiopulmonary bypass to optimize dosing regimens for children undergoing cardiac surgery. *J. Antimicrob. Chemother.* **72**, 791–800 (2016).
7. Vinks, A., Emoto, C. & Fukuda, T. Modeling and simulation in pediatric drug therapy: application of pharmacometrics to define the right dose for children. *Clin. Pharmacol. Ther.* **98**, 298–308 (2015).
8. Illamola, S.M., Colom, H. & Hasselt, J.G.C. Evaluating renal function and age as predictors of amikacin clearance in neonates: model-based analysis and optimal dosing strategies. *Br. J. Clin. Pharmacol.* **82**, 793–805 (2016).
9. van Hasselt, J.G.C. *et al*. Semiphysiological versus empirical modelling of the population pharmacokinetics of free and Total cefazolin during pregnancy. *Biomed. Res. Int.* **2014**, 1–9 (2014).
10. van Hasselt, J.G.C. *et al*. Optimizing anticancer drug treatment in pregnant cancer patients: pharmacokinetic analysis of gestation-induced changes for doxorubicin, epirubicin, docetaxel and paclitaxel. *Ann. Oncol.* **25**, 2059–2065 (2014).
11. Yoneyama, K. *et al*. A pharmacometric approach to substitute for a conventional dose-finding study in rare diseases: example of phase III dose selection for emicizumab in hemophilia A. *Clin. Pharmacokinet.* **57**, 1123–1134 (2018).
12. van Hasselt, J.G.C., van Eijkelenburg, N.K.A., Beijnen, J.H., Schellens, J.H.M. & Huitema, A.D.R. Design of a drug-drug interaction study of vincristine with azole antifungals in pediatric cancer patients using clinical trial simulation. *Pediatr. Blood Cancer* **61**, 2223–2229 (2014).
13. van Hasselt, J.G.C., Schellens, J.H.M., Beijnen, J.H. & Huitema, A.D.R. Design of informative renal impairment studies: evaluation of the impact of design stratification on bias, precision and dose adjustment error. *Invest. New Drugs* **32**, 913–927 (2014).
14. Jamei, M., Dickinson, G.L. & Rostami-Hodjegan, A. A framework for assessing inter-individual variability in pharmacokinetics using virtual human populations and integrating general knowledge of physical chemistry, biology, anatomy, physiology and genetics: a tale of 'bottom-up' vs 'top-down' recognition. *Drug Metab. Pharmacokinet.* **24**, 53–75 (2009).
15. Efron, B. Bootstrap methods: another look at the jackknife. *Ann. Stat.* **7**, 1–26 (1979).
16. Smania, G. & Jonsson, E.N. Conditional distribution modeling as an alternative method for covariates simulation: comparison with joint multivariate normal and bootstrap techniques. *CPT Pharmacometrics Syst. Pharmacol.* **10**, 330–339 (2021).
17. Tannenbaum, S.J., Holford, N.H.G., Lee, H., Peck, C.C. & Mould, D.R. Simulation of correlated continuous and categorical variables using a single multivariate distribution. *J. Pharmacokinet. Pharmacodyn.* **33**, 773–794 (2006).
18. Teutonico, D. *et al*. Generating virtual patients by multivariate and discrete re-sampling techniques. *Pharm. Res.* **32**, 3228–3237 (2015).
19. Sklar, A. Random variables, joint distribution functions, and copulas. *Kybernetika* **9**, 449–460 (1973).
20. Czado, C. Analyzing dependent data with vine copulas. *Lect. Notes Stat.* **222**, 1–92 (2019).
21. Nagler, T. & Czado, C. Evading the curse of dimensionality in nonparametric density estimation with simplified vine copulas. *J. Multivar. Anal.* **151**, 69–89 (2016).
22. De Cock, R.F.W. *et al*. Simultaneous pharmacokinetic modeling of gentamicin, tobramycin and vancomycin clearance from neonates to adults: towards a semi-physiological function for maturation in glomerular filtration. *Pharm. Res.* **31**, 2643–2654 (2014).
23. Patel, J.P., Green, B., Patel, R.K., Marsh, M.S., Davies, J.G. & Arya, R. Population pharmacokinetics of enoxaparin during the antenatal period. *Circulation* **128**, 1462–1469 (2013).
24. Johnson, A. *et al*. MIMIC-IV (version 2.0). *PhysioNet* (2022). https://doi.org/10.13026/7vcr-e114.

25. Aas, K., Czado, C., Frigessi, A. & Bakken, H. Pair-copula constructions of multiple dependence. *Insur. Math. Econ.* **44**, 182–198 (2009).
26. Nagler, T. & Vatter, T. kde1d: Univariate Kernel Density Estimation (2020) <https://cran.r-project.org/package=kde1d>.
27. Dißmann, J., Brechmann, E.C., Czado, C. & Kurowicka, D. Selecting and estimating regular vine copulae and application to financial returns. *Comput. Stat. Data Anal.* **59**, 52–69 (2013).
28. Nagler, T. & Vatter, T. rvinecopulib: High Performance Algorithms for Vine Copula Modeling (2021) <https://cran.r-project.org/package=rvinecopulib>.
29. Nagler, T. A generic approach to nonparametric function estimation with mixed data. *Stat. Probab. Lett.* **137**, 326–330 (2017).
30. Grimsley, C. & Thomson, A.H. Pharmacokinetics and dose requirements of vancomycin in neonates. *Arch. Dis. Child. - Fetal Neonatal Ed.* **81**, F221–F227 (1999).
31. Yow, H.Y., Govindaraju, K., Lim, A.H. & Abdul Rahim, N. Optimizing antimicrobial therapy by integrating multi-omics with pharmacokinetic/pharmacodynamic models and precision dosing. *Front. Pharmacol.* **13**, 1–12 (2022).
32. Conrado, D.J., Karlsson, M.O., Romero, K., Sarr, C. & Wilkins, J.J. Open innovation: towards sharing of data, models and workflows. *Eur. J. Pharm. Sci.* **109**, S65–S71 (2017).
33. Gambs, S., Ladouceur, F., Laurent, A. & Roy-Gaumond, A. Growing synthetic data through differentially-private vine copulas. *Proc. Priv. Enhancing Technol.* **2021**, 122–141 (2021).
34. Czado, C. & Nagler, T. Vine copula based modeling. *Annu. Rev. Stat. Its Appl.* **9**, 453–477 (2022).
35. Faugeras, O.P. Inference for copula modeling of discrete data: a cautionary tale and some facts. *Depend. Model.* **5**, 121–132 (2017).
36. Nguyen, T.H.T. *et al*. Model evaluation of continuous data pharmacometric models: metrics and graphics. *CPT Pharmacometrics Syst. Pharmacol.* **6**, 87–109 (2017).