Statistical learning for complex data to enable precision medicine strategies

Zwep, L.B.

Chapter 7

Virtual patient simulation using copula modeling

Authors

Laura B. Zwep

Tingjie Guo

Thomas Nagler

Catherijne A. J. Knibbe

Jacqueline J. Meulman

J. G. Coen van Hasselt

Abstract

Clinical trial design and dosing optimization strategies are increasingly relying on model-based approaches in pharmacometrics and quantitative systems pharmacology (QSP), which incorporate patient characteristics to simulate the expected pharmacokinetic (PK) or pharmacodynamic (PD) response in cohorts of virtual patients. To this end, the individual-level patient characteristics, or covariates, are used as input for such simulations should accurately reflect the values seen in real patient populations. Current methods to achieve this goal either 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. Here, 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 to a number of case studies. Our analyses demonstrate 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.

7.1 Introduction

Model-based approaches in pharmacometrics and quantitative systems pharmacology (QSP) (Bonate, 2000, 2001; Chelliah et al., 2020) have become of pivotal importance for the optimization of drug treatment strategies or clinical trial designs (Holford et al., 2010; Langenhorst et al., 2020). These model-based approaches typically simulate the expected pharmacokinetic (PK) and/or pharmacodynamic (PD) response and the associated inter-individual variability for a cohort of virtual patients. Here, the inter-individual 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 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 (De Cock et al., 2016), such as pediatric (Illamola et al., 2016; J. G. C. van Hasselt, Allegaert, et al., 2014; Vinks et al., 2015) or pregnant patients (J. G. van Hasselt et al., 2014), or, to evaluate different potential trial designs in specific types of

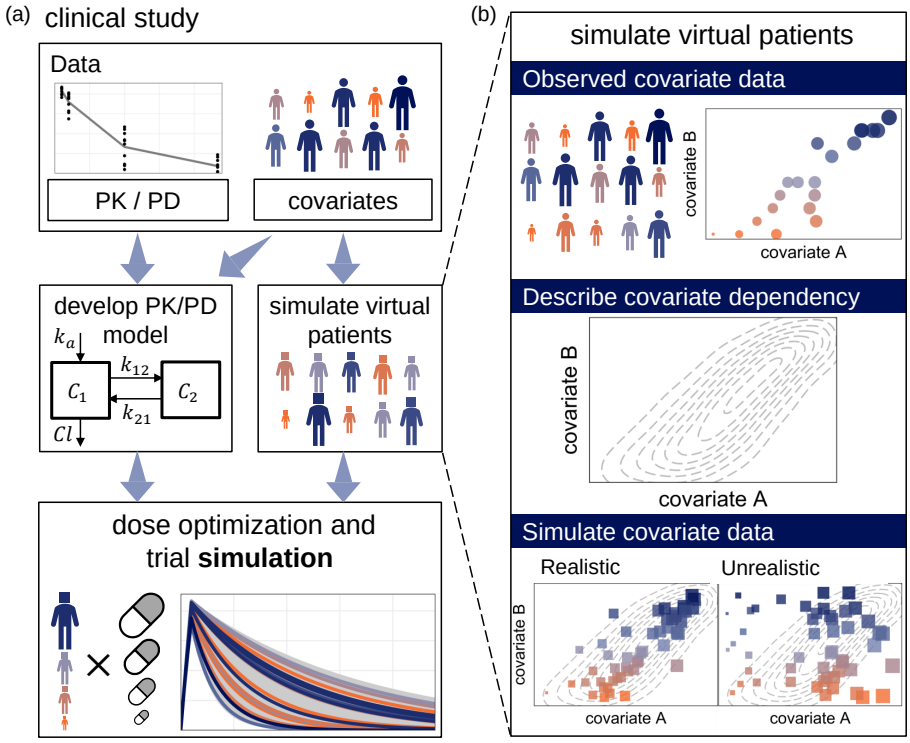

Figure 7.1: Pharmacometric workflow. (a) In order to optimize dosing for new medication or special patient populations, pharmacometric models, such as (PK/)PD models, are used to simulate new patient dosing regimens. Next to the developed pharmacometric model, simulation studies require covariate simulation. (b) An important challenge for covariate simulation is sampling realistic patients, where the dependency between covariates is preserved.

patients or treatments(J. G. C. van Hasselt, Allegaert, et al., 2014; J. G. C. van Hasselt, van Eijkelenburg, et al., 2014; Yoneyama et al., 2017) (Figure 7.1a).

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, sex, age), organ function measures (e.g. renal or hepatic function), pharmacodynamic endpoints (cardiovascular readouts, 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 7.1b). 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. 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 (Efron, 1979), 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 (CD) (Smania & Jonsson, 2021), 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 (MVND) (Tannenbaum et al., 2006) do capture the correlation structure (Teutonico et al., 2015), 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 shortcomings of alternative distribution-based methods while not requiring access to patient-level data (Czado, 2019; Nagler & Czado, 2016; Sklar, 1973). 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 pharmacokinetic simulations, time-varying covariates, and higher-dimensional covariates.

7.2 Methods

7.2.1 Data

Three different datasets of combined patient characteristics were used in this study to evaluate the performance and explore different applications. The first data set

contains a special patient population of pediatric patients (Cock et al., 2014) with 445 neonates and young children admitted to the ICU, with twelve measured covariates, including body weight, serum creatinine level (SCr) and age. These data were used to evaluate the simulation performance (Data set 1). A second data set on pregnancy data (Patel et al., 2013) with 123 subjects, with biomarkers measured over time, was used to simulate longitudinal covariate profiles (Data set 2). Lastly, MIMIC (Johnson et al., 2022), a large observational dataset with ICU patients, was used to evaluate the correlation structure between a large set of 30 variables for >53,000 patients (Data set 3).

7.2.2 Copula estimation and simulation

Vine copulas were used to estimate the joint density between all covariates. 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 (Nagler & Vatter, 2020). 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. For each bivariate copula, a set of parametric distributions was fit and the best fitting distributions were chosen by minimizing the AIC. Vine copulas with different distributions were fit using the R library rvinecopulib (Nagler & Vatter, 2021). 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 (`https://github.com/vanhasseltlab/copula_vps`).

7.2.3 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 (Cock et al., 2014) (Data set 1). The estimation and simulation were performed in two differently sized covariate sets, with the same subjects, but a different number of covariates: one simulation on three covariates, age, SCr and body weight, and one on twelve covariates. The distribution of the simulated population was compared with the distribution of the observed population in terms of the mean and standard deviation for each covariate and 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 to four other simulation methods, of which two method are based on patient-level data and two methods are based on characteriza-

tion of the covariate distribution. Bootstrap simulations were conducted by resampling full rows from the original data with replacement (Efron, 1979). The conditional distribution (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 (Smania & Jonsson, 2021). The standard multiple imputation method 'predictive mean matching' was used, corresponding to their paper. The distribution-based methods used were the multivariate normal distribution (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 (Nagler, 2017; Nagler & Vatter, 2020). Covariate values were sampled from the respective density functions.

7.2.4 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 performance of the use of body weight and SCr from the three-covariate copula and the MDs simulation was compared in a population PK one-compartmental model for vancomycin (Grimsley & Thomson, 1999).

$$\frac{dA}{dt} = 0 - \frac{Cl}{V} \cdot A$$

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

$$V = 0.669 \cdot WT$$

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 MDs simulations. These PK profiles were compared using the AUC of the first 24 hours after dosing. The correlation between the AUC and the covariates, 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 six time-varying covariates (y) over the gestational age (t) during pregnancy (Patel et al., 2013), 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 three individual parameters for each patient i, the intercept ($\beta_{0j} + b_{0ji}$),

the linear term ($\beta_{1j} + b_{1ji}$) and the quadratic term ($\beta_{2j} + b_{2ji}$), resulting in a total of 18 dimensions.

$$\hat{y}_{ij}(t) = \beta_{0j} + b_{0ji} + \beta_{1j} \cdot t + b_{1ji} \cdot t + \beta_{2j} \cdot t^2 + b_{2ji} \cdot t^2$$
$$b_{0ji} \sim N(0, \sigma_0)$$
$$b_{1ji} \sim N(0, \sigma_1)$$
$$b_{2ji} \sim N(0, \sigma_2)$$

For example, yielding for albumin concentration:

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

$$b_{0ji} \sim N(0, 1.86)$$
$$b_{1ji} \sim N(0, 0.105)$$
$$b_{2ji} \sim N(0, 0.00224)$$

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 dataset, copula simulation was used to characterize and simulate from the MIMIC database (Data set 3) (Johnson et al., 2022). A copula model was fit to a large dataset 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.

7.3 Results

7.3.1 Evaluation of simulation performance

The performance of the copulas was assessed on two differently sized datasets, one with three covariates and one with twelve covariates (Data set 1). First, for a set of three covariates, copulas show a low relative error of 0.02, 0.08 and 0.04 for the in terms of correlations between age and body weight, age and SCr, and body weight and SCr respectively (Figure 7.2a). Second, for the twelve-covariate simulations, the copula simulation slightly underestimates the covariances with a median error of 0.05 over all covariate combinations (Figure 7.2b).

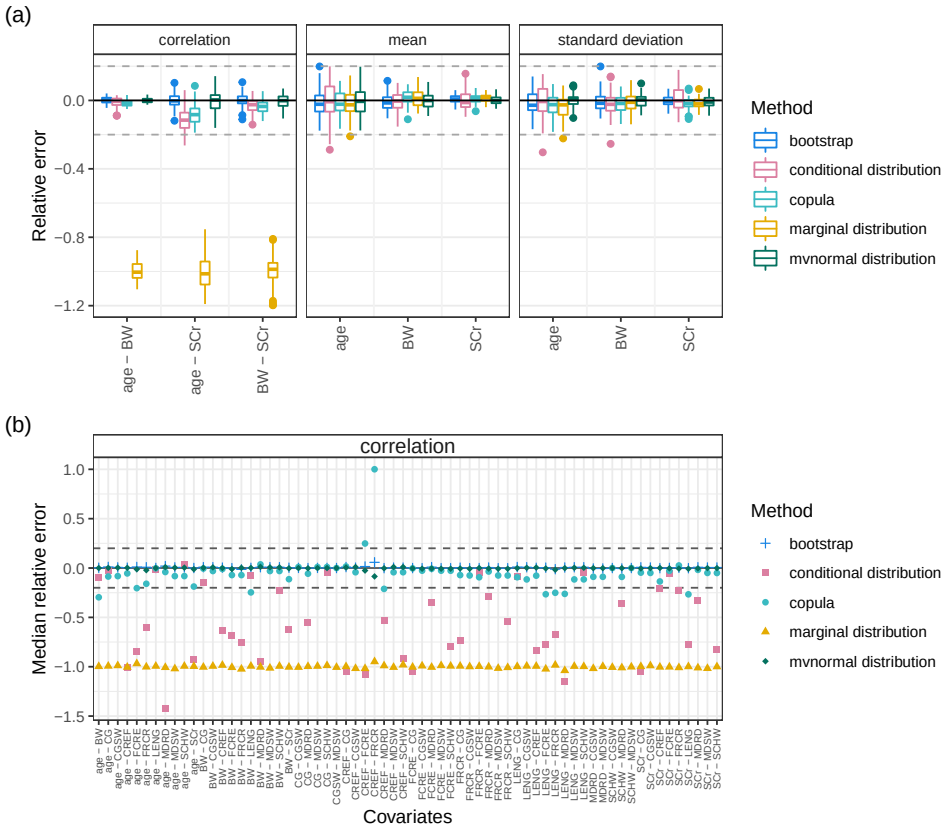

Figure 7.2: Relative error over 100 simulations as compared to the statistics of the observed population for five different simulation methods. (a) Boxplots of the correlation, mean and standard deviation of three covariates. (b) Median relative error of a large covariate simulation for the correlations of each combination of twelve covariates.

The performance of copulas was compared to four other simulation methods. For the three-covariate simulation the copula yielded similar results to the conditional distributions, which has relative errors of 0.01, 0.12 and 0.03 (Figure 7.2a), but for the

twelve-covariate simulations, the CD simulations show a large median underestimation with a relative error of 0.60 (Figure 7.2b). The bootstrap shows the best performance, since it can fully keep the dependence structure intact, both in the three-covariate (Figure 7.2a) and the twelve-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 S7.1).

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 twelve-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 S7.1).

7.3.2 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 7.3a). 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 MDs simulations (r = -0.07), whereas the copula preserved their dependence (r = 0.66) (Figure 7.3b). 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

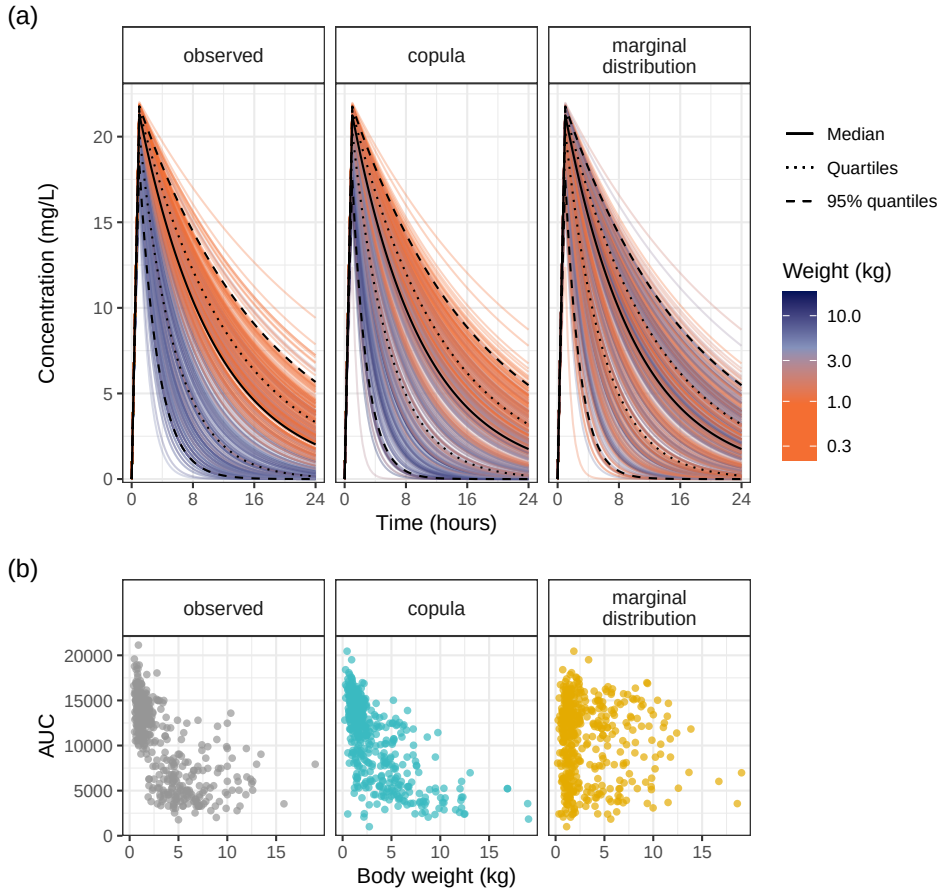

Figure 7.3: (a) Pharmacokinetic (PK) curves calculated for the observed population and the virtual patient populations from copula simulations and the marginal densities (MDs). The median and quantiles show a similar pattern between all three sets, however the weight is randomly distributed over the PK profiles for the simulation with MDs. (b) Scatter plot of area under the PK curve (AUC) against body weight.

linear and quadratic terms. Polynomial regression coefficients were simulated in a realistic domain, while simulating from a MDs led to more extreme polynomial curves, with a five times higher error on the standard deviation of the AUC (Figure 7.4). 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 dataset, 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

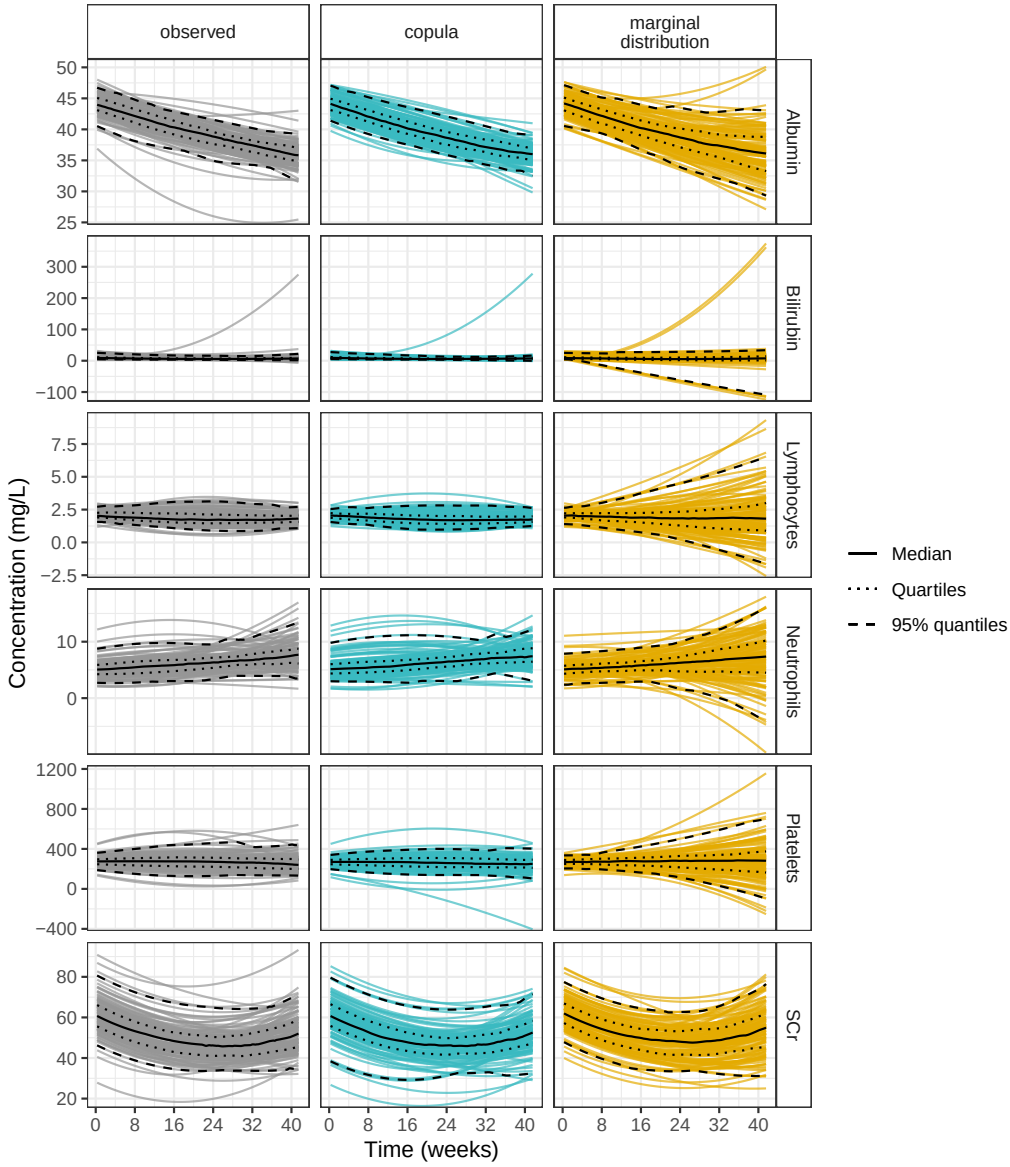

Figure 7.4: Polynomial curves for the six biomarkers from pregnancy data. In gray the estimated curves from the observed data. The copula (turquoise) shows very similar patterns, while the marginal distribution (yellow) shows extreme values, especially at the end of the curve.

0.11, which was slightly worse compared to the estimation in the lower dimensional twelve- and three-covariate data sets. Some covariates show interesting dependency structures, which can be evaluated and be used in covariate selection decision making (Figure 7.5). The results from the larger data set also show that through the use of copulas, it is feasible to share hospital data distributions.

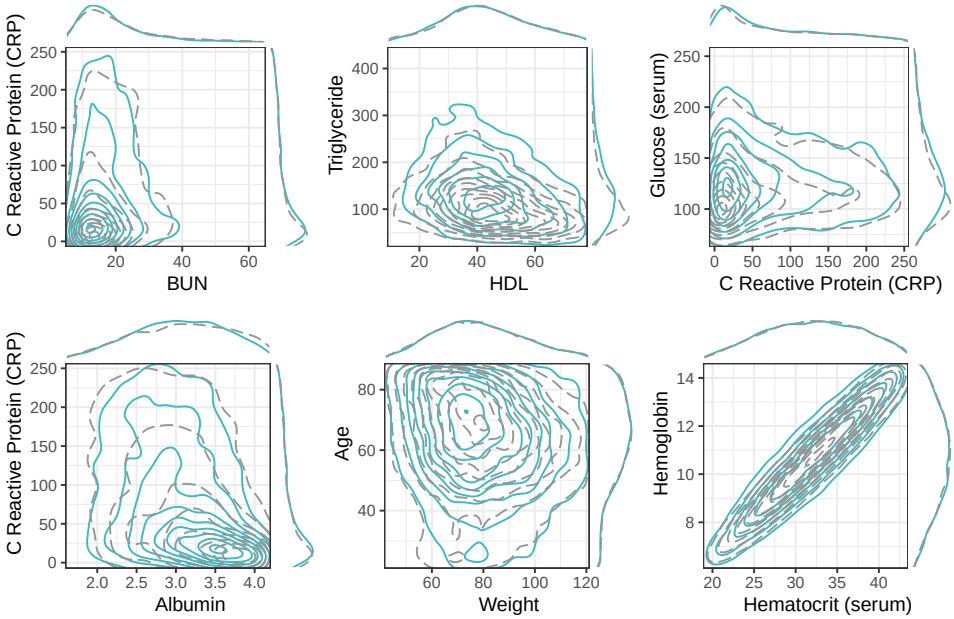

Figure 7.5: Set of selected covariate combinations with the densities of the observed population (gray dashed line) and the simulated population from a copula (blue solid lines), with marginal densities on the top and right sides of each plot. More overlap between the lines shows a better correspondence between the observed and simulated patient covariates.

7.4 Discussion

We showed a competitive or superior performance of copula simulations compared to 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 datasets. Preserving the dependence structure in copula simulations allows for simulating covariate sets for realistic PK predictions, time-varying covariates, and in a large scale data set, i.e., 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, since these either neglect the dependence structure among the covariates, or rely on real patient data in simulation.

We evaluated the performance of copulas compared to other simulation methods. While performing well in lower dimensions, we observed increasing underestimation in higher dimensions for CD, making the method less suitable for simulations in higher dimension. The MVND showed very promising results in terms of capturing the correlation (Figure 7.2). 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 S7.1). 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. 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 AUC, which is of high clinical relevance. This feature of copulas provides a significant insight into how PK may differ between subgroups of patients. It allows one to optimize the dose for a particular patient group or to study the differences between patients groups. We found that PK at the population level is not affected by the method used for virtual patient simulation (Figure 7.3). 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 level. 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 (Conrado et al., 2017). 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 (Gambs et al., 2021). Using copulas both opens opportunities for better replication and comparison studies, and copulas can facilitate in simulation platforms for sharing patient characteristics. 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 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 7.6).

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 (Czado & Nagler, 2022), however the copula method is not able to deal with unordered categorical variables in a natural way (Faugeras, 2017).

Regardless of the method of simulation, further research would also require looking into the underestimation of the correlation by the different simulation techniques, since 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

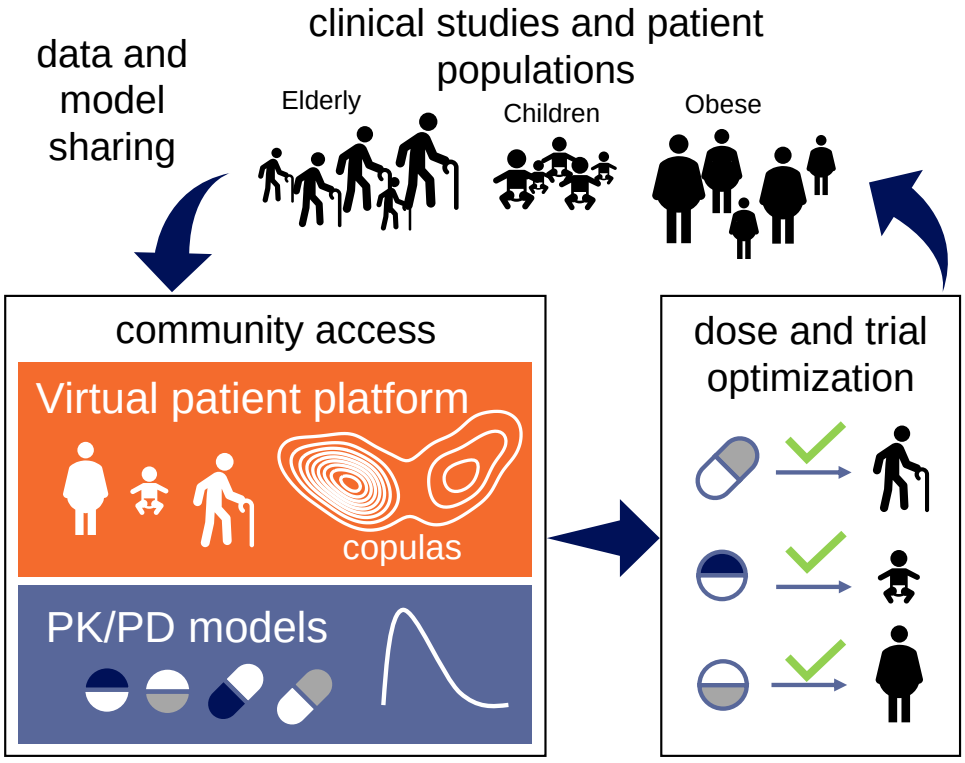

Figure 7.6: Community access pharmacometrics research pipeline. Data and pharmacometric models from (special) patient populations can be shared with the clinical pharmacology community. Through copulas, covariate sets can be simulated, which, when used in PK/PD models, can aid treatment and dosing optimization, ultimately improving treatment for the patients.

looking at the outcomes of pharmacometric models (Nguyen et al., 2017).

In summary, copulas represent an attractive approach to capture multivariate covariate distributions, which can be readily implemented for pharmacometric simulations. Importantly, the distribution-based nature of copula's has the distinct advantage that access to original individual-level datasets 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.

References

Bonate, P. L. (2000). Clinical trial simulation in drug development. *Pharmaceutical Research*, *17*(3), 252–256. doi: 10.1023/A:1007548719885

Bonate, P. L. (2001). A Brief Introduction to Monte Carlo Simulation. *Clinical Pharmacokinetics*, *40*(1), 15–22. Retrieved from http://link.springer.com/10.2165/00003088-200140010-00002 doi: 10.2165/00003088-200140010-00002

Chelliah, V., Lazarou, G., Bhatnagar, S., Gibbs, J. P., Nijsen, M., Ray, A., ... Kierzek, A. M. (2020, Aug). Quantitative systems pharmacology approaches for immuno-oncology: Adding virtual patients to the development paradigm. *Clinical Pharmacology & Therapeutics*, *109*(3), 605–618. Retrieved from https://doi.org/10.1002%2Fcpt.1987 doi: 10.1002/cpt.1987

Cock, R. F. W. D., Allegaert, K., Brussee, J. M., Sherwin, C. M. T., Mulla, H., de Hoog, M., ... Knibbe, C. A. J. (2014, May). Simultaneous pharmacokinetic modeling of gentamicin, tobramycin and vancomycin clearance from neonates to adults: Towards a semi-physiological function for maturation in glomerular filtration. *Pharmaceutical Research*, *31*(10), 2643–2654. Retrieved from https://doi.org/10.1007%2Fs11095-014-1361-z doi: 10.1007/s11095-014-1361-z

Conrado, D. J., Karlsson, M. O., Romero, K., Sarr, C., & Wilkins, J. J. (2017, Nov). Open innovation: Towards sharing of data, models and workflows. *European Journal of Pharmaceutical Sciences*, *109*, S65–S71. Retrieved from https://doi.org/10.1016%2Fj.ejps.2017.06.035 doi: 10.1016/j.ejps.2017.06.035

Czado, C. (2019). Simulating regular vine copulas and distributions. In *Analyzing dependent data with vine copulas* (pp. 123–144). Springer International Publishing. Retrieved from https://doi.org/10.1007%2F978-3-030-13785-4_6 doi: 10.1007/978-3-030-13785-4_6

Czado, C., & Nagler, T. (2022, Mar). Vine Copula Based Modeling. *Annual Review of Statistics and Its Application*, *9*(1), 453–477. Retrieved from https://www.annualreviews.org/doi/10.1146/annurev-statistics-040220-101153 doi: 10.1146/annurev-statistics-040220-101153

De Cock, P. A. J. G., Mulla, H., Desmet, S., De Somer, F., McWhinney, B. C., Ungerer, J. P. J., ... De Paepe, P. (2016, Dec). Population pharmacokinetics of cefazolin before, during and after cardiopulmonary bypass to optimize dosing regimens for children undergoing cardiac surgery. *Journal of Antimicrobial Chemotherapy*, *72*(3), dkw496. Retrieved from https://academic.oup.com/jac/article-lookup/doi/10.1093/jac/dkw496 doi: 10.1093/jac/dkw496

Efron, B. (1979, Jan). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, *7*(1). Retrieved from https://doi.org/10.1214%2Faos%2F1176344552 doi: 10.1214/aos/1176344552

Faugeras, O. P. (2017, Jan). Inference for copula modeling of discrete data: a cautionary tale and some facts. *Dependence Modeling*, *5*(1), 121–132. Retrieved from https://www.degruyter.com/document/doi/10.1515/demo-2017-0008/html doi: 10.1515/demo-2017-0008

Gambs, S., Ladouceur, F., Laurent, A., & Roy-Gaumond, A. (2021, Apr). Growing synthetic data through differentially-private vine copulas. *Proceedings on Privacy Enhancing Technologies*, *2021*(3), 122–141. Retrieved from https://doi.org/10.2478%2Fpopets-2021-0040 doi: 10.2478/popets-2021-0040

Grimsley, C., & Thomson, A. H. (1999, Nov). Pharmacokinetics and dose requirements of vancomycin in neonates. *Archives of Disease in Childhood - Fetal and Neonatal Edition*, *81*(3), F221–F227. Retrieved from https://doi.org/10.1136%2Ffn.81.3.f221 doi: 10.1136/fn.81.3.f221

Holford, N., Ma, S. C., & Ploeger, B. A. (2010, Jul). Clinical trial simulation: A review. *Clinical Pharmacology & Therapeutics*, *88*(2), 166–182. Retrieved from https://doi.org/10.1038%2Fclpt.2010.114 doi: 10.1038/clpt.2010.114

Illamola, S. M., Colom, H., & Hasselt, J. G. C. (2016, Sep). Evaluating renal function and age as predictors of amikacin clearance in neonates: model□based analysis and optimal dosing strategies. *British Journal of Clinical Pharmacology*, *82*(3), 793–805. Retrieved from https://onlinelibrary.wiley.com/doi/10.1111/bcp.13016 doi: 10.1111/bcp.13016

Johnson, A., Bulgarelli, L., Pollard, T., Horng, S., Celi, L. A., & Mark, R. (2022). *MIMIC-IV*. PhysioNet. Retrieved from https://physionet.org/content/mimiciv/1.0/ doi: 10.13026/7vcr-e114

Langenhorst, J. B., Dorlo, T. P., Kesteren, C., Maarseveen, E. M., Nierkens, S., Witte, M. A., ... Huitema, A. D. (2020, Apr). Clinical trial simulation to optimize trial design for fludarabine dosing strategies in allogeneic hematopoietic cell transplantation. *CPT: Pharmacometrics & Systems Pharmacology*. Retrieved from https://doi.org/10.1002%2Fpsp4.12486 doi: 10.1002/psp4.12486

Nagler, T. (2017, May). Asymptotic analysis of the jittering kernel density estimator. *Mathematical Methods of Statistics*, *27*(1), 32–46. Retrieved from http://arxiv.org/abs/1705.05431 doi: 10.3103/S1066530718010027

Nagler, T., & Czado, C. (2016, Oct). Evading the curse of dimensionality in nonparametric density estimation with simplified vine copulas. *Journal of Multivariate Analysis*, *151*, 69–89. Retrieved from https://doi.org/10.1016%2Fj.jmva.2016.07.003 doi: 10.1016/j.jmva.2016.07.003

Nagler, T., & Vatter, T. (2020). kde1d: Univariate Kernel Density Estimation [Computer software manual]. Retrieved from https://cran.r-project.org/package=kde1d

Nagler, T., & Vatter, T. (2021). rvinecopulib: High Performance Algorithms for Vine Copula Modeling [Computer software manual]. Retrieved from https://cran.r-project.org/package=rvinecopulib

Nguyen, T. H., Mouksassi, M. S., Holford, N., Al-Huniti, N., Freedman, I., Hooker, A. C., ... Mentre, F. (2017). Model evaluation of continuous data pharmacometric models: Metrics and graphics. *CPT: Pharmacometrics and Systems Pharmacology*, *6*(2), 87–109. doi: 10.1002/psp4.12161

Patel, J. P., Green, B., Patel, R. K., Marsh, M. S., Davies, J. G., & Arya, R. (2013, Sep). Population pharmacokinetics of enoxaparin during the antenatal period. *Circulation*, *128*(13), 1462–1469. Retrieved from https://doi.org/10.1161%2Fcirculationaha.113.003198 doi: 10.1161/circulationaha.113.003198

Sklar, A. (1973). Random variables, joint distribution functions, and copulas. *Kybernetika*, *9*(6), 449–460.

Smania, G., & Jonsson, E. N. (2021, Apr). Conditional distribution modeling as an alternative method for covariates simulation: Comparison with joint multivariate normal and bootstrap techniques. *CPT: Pharmacometrics & Systems Pharmacology*, *10*(4), 330–339. Retrieved from https://doi.org/10.1002%2Fpsp4.12613 doi: 10.1002/psp4.12613

Tannenbaum, S. J., Holford, N. H. G., Lee, H., Peck, C. C., & Mould, D. R. (2006, Oct). Simulation of correlated continuous and categorical variables using a single multivariate distribution. *Journal of Pharmacokinetics and Pharmacodynamics*, *33*(6), 773–794. Retrieved from https://doi.org/10.1007%2Fs10928-006-9033-1 doi: 10.1007/s10928-006-9033-1

Teutonico, D., Musuamba, F., Maas, H. J., Facius, A., Yang, S., Danhof, M., & Pasqua, O. D. (2015, May). Generating virtual patients by multivariate and discrete re-sampling techniques. *Pharmaceutical Research*, *32*(10), 3228–3237. Retrieved from https://doi.org/10.1007%2Fs11095-015-1699-x doi: 10.1007/s11095-015-1699-x

van Hasselt, J. G., van Calsteren, K., Heyns, L., Han, S., Mhallem Gziri, M., Schellens, J. H., ... Amant, F. (2014). Optimizing anticancer drug treatment in pregnant cancer patients: pharmacokinetic analysis of gestation-induced changes for doxorubicin, epirubicin, docetaxel and paclitaxel. *Annals of Oncology*, *25*(10), 2059–2065. Retrieved from https://doi.org/10.1093/annonc/mdu140 doi: 10.1093/annonc/mdu140

van Hasselt, J. G. C., Allegaert, K., van Calsteren, K., Beijnen, J. H., Schellens, J. H. M., & Huitema, A. D. R. (2014). Semiphysiological versus Empirical Modelling of the Population Pharmacokinetics of Free and Total Cefazolin during Pregnancy. *BioMed Research International*, *2014*, 1–9. Retrieved from http://www.hindawi.com/journals/bmri/2014/897216/ doi: 10.1155/2014/897216

van Hasselt, J. G. C., van Eijkelenburg, N. K. A., Beijnen, J. H., Schellens, J. H., & Huitema, A. D. R. (2014, Dec). Design of a drug-drug interaction study of vincristine with azole antifungals in pediatric cancer patients using clinical trial simulation. *Pediatric Blood & Cancer*, *61*(12), 2223–2229. Retrieved from https://onlinelibrary.wiley.com/doi/10.1002/pbc.25198 doi: 10.1002/pbc.25198

Vinks, A., Emoto, C., & Fukuda, T. (2015, Jul). Modeling and simulation in pediatric drug therapy: Application of pharmacometrics to define the right dose for children. *Clinical Pharmacology & Therapeutics*, *98*(3), 298–308. Retrieved from https://doi.org/10.1002%2Fcpt.169 doi: 10.1002/cpt.169

Yoneyama, K., Schmitt, C., Kotani, N., Levy, G. G., Kasai, R., Iida, S., ... Kawanishi, T. (2017, Dec). 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. *Clinical Pharmacokinetics*, *57*(9), 1123–1134. Retrieved from https://doi.org/10.1007%2Fs40262-017-0616-3 doi: 10.1007/s40262-017-0616-3

Supplementary material

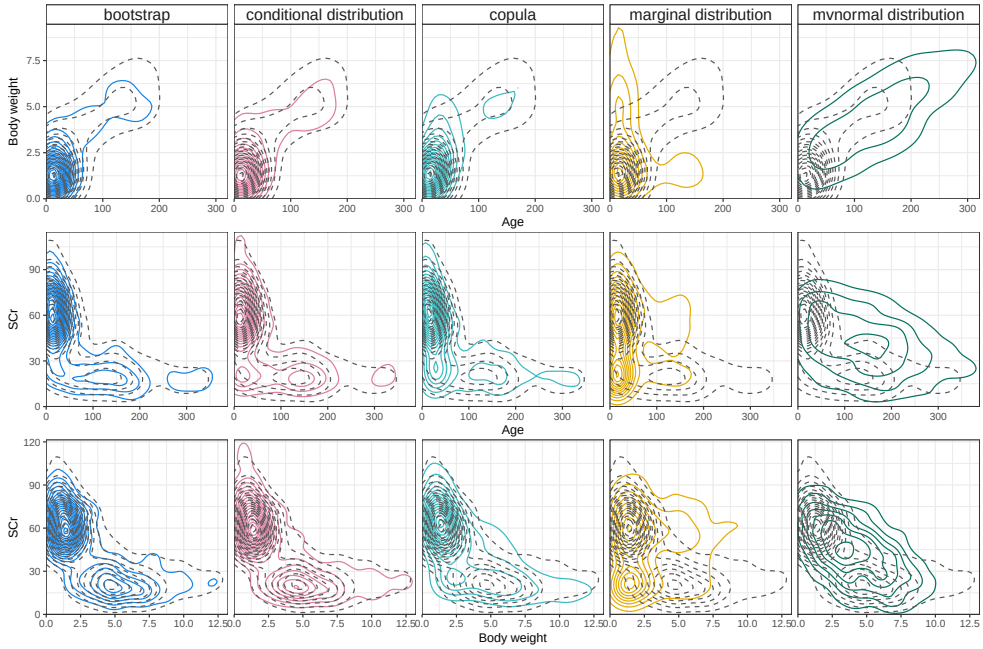

Figure S7.1: Densities of the three covariate simulations. Grey dashed lines show the observed joint density for each pair of covariates. The solid lines represent the joint density of a simulated population for each of the five simulation methods: bootstrap (blue), conditional distributions (pink), copula (turquoise), marginal distribution (yellow) and multivariate normal distribution (green).