Constructing a representative in-silico population for paediatric simulations: Application to HIV-positive African children

Roeland E. Wasmann\textsuperscript{1} | Elin M. Svensson\textsuperscript{2,3} | A. Sarah Walker\textsuperscript{4} | Michelle N. Clements\textsuperscript{4} | Paolo Denti\textsuperscript{1}

\textsuperscript{1}Division of Clinical Pharmacology, Department of Medicine, University of Cape Town, Cape Town, South Africa
\textsuperscript{2}Department of Pharmaceutical Biosciences, Uppsala University, Uppsala, Sweden
\textsuperscript{3}Department of Pharmacy, Radboud Institute for Health Sciences, Radboud University Medical Center, Nijmegen, The Netherlands
\textsuperscript{4}MRC Clinical Trials Unit at UCL, London, UK

Correspondence
Elin M. Svensson, Radboud University Medical Center, PO Box 9101, 6500 HB. Nijmegen (864), The Netherlands.
Email: elin.svensson@radboudumc.nl

Aims: Simulations are an essential tool for investigating scenarios in pharmacokinetics-pharmacodynamics. The models used during simulation often include the effect of highly correlated covariates such as weight, height and sex, and for children also age, which complicates the construction of an in silico population. For this reason, a suitable and representative patient population is crucial for the simulations to produce meaningful results. For simulation in paediatric patients, international growth charts from the World Health Organization (WHO) and the Centers for Disease Control and Prevention (CDC) provide a reference, but these may not always be representative for specific populations, such as malnourished children with HIV or acutely unwell children.

Methods: We present a workflow to construct a virtual paediatric patient population using WHO and CDC growth charts, suggest piecewise linear functions to adjust the median of the growth charts by sex and age, and suggest visual diagnostics to compare with the target population. We applied this workflow in a population of 1206 HIV-positive African children, consisting of 19 742 observations with weight ranging from 3.8 to 79.7 kg, height from 55.5 to 180 cm, and an age between 0.40 and 18 years.

Results: Before adjustment, the WHO and CDC charts produced weights and heights higher compared to the observed data. After applying our methodology, we could simulate weight, height, sex and age combinations in good agreement with the observed data.

Conclusion: The methodology presented here is flexible and may be applied to other scenarios where WHO and CDC growth standards might not be appropriate. In addition we provide R scripts and a large ready-to-use paediatric population.

KEYWORDS
modelling, paediatric population, pharmacokinetics-pharmacodynamics, simulation, underweight, weight-for-age
Modelling and simulation are increasingly important within pharmacokinetics-pharmacodynamics to investigate alternative scenarios or support decision making in trial design or public health policy. Besides a model and a relevant outcome (eg, exposure metrics, survival), a realistic target population is required. This means that the covariate values reflect those in the intended use population and that they maintain their respective correlation (eg, weight and height).

Two widely used approaches to accomplish this are (a) nonparametric bootstrapping of historical data (ie, re-sampling from a real population) and (b) simulating an entirely new population using parametric distributions for relevant covariate values. However, both approaches require a (relatively large) representative population from a real population) and (b) simulating an entirely new population (a) nonparametric bootstrapping of historical data (ie, re-sampling

What this study adds

- A flexible methodology to generate a representative in silico paediatric population can aid pharmacometricians in making better predictions in vulnerable patient populations.
Observed patient data

Data from HIV-positive children included in the ARROW trial was available and used to (a) assess the natural correlation between z-scores of weight and height (i.e., a child that is underweight is more likely to also be shorter than average) using equation 1, (b) compare with the simulated population and (c) adjust the simulated population with a sex- and age-dependent function.9 A second cohort of HIV-positive children below 25 kg from the CHAPAS-3 trial was used to validate the final virtual patient population.10

Adjusted simulated population

The data generated from the reference growth charts was graphically compared with the observed data. To determine a plausible correlation between weight and height in the population of interest, the individual z-scores weight-for-age and height-for-age were calculated on the observed data using eq. 1, and their correlation was determined. This correlation was used to generate multivariate normally distributed z-scores for weight and height where \( Z \sim N(0, 1) \).

To adjust for stunting and wasting we first calculated the difference between the observed height and weight and the median (\( M_{\text{s,a}} \)-value provided by the growth charts) by dividing the observed data points by the \( M_{\text{s,a}} \)-value to obtain an adjustment factor. Piecewise linear functions were used to model this adjustment factor over sex and age to obtain a correction function, \( f \) (sex, age), basing model selection on the Bayesian Information Criterion (BIC).

The multivariate normally distributed z-scores, together with the LMS parameters from the growth charts, were used to generate a simulated population using eq. 4:

\[
X_{\text{s,a}} = f(\text{sex, age}) \times M_{\text{s,a}} \times (1 + L_{\text{s,a}} \times S_{\text{s,a}} \times Z)^{1/L_{\text{s,a}}}
\]

The weight, height and FFM of the simulated and observed population were graphically compared again. Finally, the simulated population was compared with the CHAPAS-3 dataset for validation. For this final comparison, simulated and observed patients weighing above 25 kg and above 12 years of age were excluded to emulate the inclusion criteria of CHAPAS-3.

All analyses were performed in R (version 4.0.0) with R Studio interface (version 1.2.5042) using tidyverse, data.table, gridExtra and ggpubr packages. Piecewise linear functions were derived in NONMEM (version 7.4.2; Icon Development Solutions, Ellicott City, MD, USA) and Perl-Speaks-NONMEM (version 4.8.8) with the Pirana (version 2.9.9) interface.11

3 | RESULTS

3.1 | Growth charts

The United States CDC growth charts generally contain a higher median weight combined with a higher variance and more skewness.

\[ X_{\text{s,a}} = M_{\text{s,a}} \times (1 + L_{\text{s,a}} \times S_{\text{s,a}} \times Z)^{1/L_{\text{s,a}}} \]
towards high weights than the WHO growth charts. We therefore adjusted the LMS parameters of the CDC charts to bring them in line with the WHO charts at 10 years. For males the parameters were adjusted downwards by 43%, 3% and 9%, and for females by 29%, 4% and 12% for the $L$ (Box-Cox transformation), $M$ (median) and $S$ (coefficient of variation) parameters, respectively (Supporting Information Figure S1).

### 3.2 | Observed patient data

The ARROW dataset contained 52,193 matched observations of height and weight in 1,206 HIV-positive children from Uganda and Zimbabwe who were followed up for up to 5 years. We included maximally one measurement every third month for each patient, thus yielding 19,742 observations. The patients had a median (range) age

![Figure 2](image-url)
of 8.0 (0.40-18) years, weight of 21.0 (3.80-79.7) kg, height of 117 (55.5-180) cm, weight-for-age z-score of −1.3 (−7.1-4.2) and height-for-age z-score of −1.9 (−8.4-3.4); 49% were female. Eighty-nine per cent of the patients had a weight-for-age z-score below 0 and 95% had a height-for-age z-score below 0, contrasting with 50% that would be expected if the growth charts reflected this population. The correlation between the z-scores for weight-for-age and height-for-age was 0.70 (Figure 1). During the simulation step this correlation was used to generate multivariate normally distributed z-scores for weight and height.

The CHAPAS-3 dataset contained 11,221 height and weight observations in 478 HIV-positive children from Uganda, Zambia and Zimbabwe. Using maximally one measurement every third month for each patient resulted in 4943 measurements with a median (range) age of 4.7 (0.22-17) years, weight of 16 (4.5-42.9) kg, height of 101 (58.7-156) cm, weight-for-age z-score of −0.89 (−6-2.1) and height-for-age z-score of −1.8 (−7.9-2.6); 51% were female. Eighty-three per cent of the patients had a weight-for-age z-score below 0 and 93% had a height-for-age z-score below 0. The correlation between the z-scores for weight-for-age and height-for-age was 0.72.

### 3.3 | Adjusted simulated population

Using the growth charts a virtual population was simulated and visually compared with the ARROW dataset (Figure 2, left column). The median, 5th and 95th percentiles for the weight, height and FFM of the simulated population are higher than those from the observed data in ARROW, indicating that children in this study had lower weights, heights and FFMs than the typical values.

The adjustment factor for weight was best described using a piecewise linear function with five breakpoints, two of which are sex-dependent (Figure 3 and Supporting Information Table S1). The adjustment factor for height was best described by a piecewise linear function with seven breakpoints, three of which are sex-dependent (Figure 3 and Supporting Information Table S2).

After adjustment, there was generally good agreement between the observed and simulated medians for weight, height and FFM (Figure 2, right column). More detailed density plots per age group (eg, infants, toddlers and adolescents) of the observed and simulated data before and after adjustment confirm a good agreement in all groups after adjustment (Supporting Information Figure S3). The external validation with the CHAPAS-3 data also showed good agreement between observed and simulated patients after adjustment of the growth charts (Figure 4).

An in silico paediatric population constructed with the final model is available in the Supporting Information. It consists of over 86,000 African HIV-positive children aged 0 to 18 years (50% female) with 200 patients per age-month. Additionally, the R-script is provided to ease the comparison of other populations and the one provided. Finally, the R-script also allows the input of an alternative correction function or simulation of a new population.

**FIGURE 3** Adjustment factors for patient weight (top) and height (bottom). The lines represent the piecewise linear functions describing the relation between age and the adjustment factor for females (red) and males (blue). The functions are given in Supporting Information Tables S1 and S2. The dashed red line represents the situation where a population is already in line with WHO and CDC growth charts.
4 | DISCUSSION

Paediatric patients are a difficult population to treat because maturation and growth need to be accounted for, both at the start of treatment and during prolonged treatment (eg, tuberculosis and HIV). This population is a moving target due to maturation and growth while children are aging. Malnutrition in some of the populations is not uncommon and these children make some of the most vulnerable patients. Because body size, whether it is weight or fat-free mass, has a large impact on pharmacokinetic parameters it is crucial that simulations that guide drug development take malnourishment into account and make predictions in an accurate target population.

However, generating a virtual population for use in modelling and simulation with multiple correlated variables can be challenging. Here, we provide a straightforward approach to construct a population using freely available data sources and software, focused on

**FIGURE 4** Agreement between the observed data from the CHAPAS-3 trial (red dots) and simulated (black dots) distribution of weight (top row), height (middle row) and fat-free mass (bottom row) before (left column) and after (right column) adjusting the simulated data using the sex- and age-dependent adjustment factor. The lines represent the median of the observed (red) and simulated (black) data.
anthropometric measures. Furthermore, we applied this method to a population of HIV-positive children from Sub-Saharan Africa and showed that our in silico population is a good representation of the real dataset, making it suitable for simulations. We provide a ready-to-use dataset containing children from birth to 18 years old representative for African HIV-infected children.

When generating and adjusting a simulated population one must carefully consider its proposed use. In this analysis we showed an example aimed at reflecting an HIV-positive paediatric population that initiates treatment and is then followed up. Therefore, we chose to use one measurement per individual every 3 months, thus reflecting children during chronic treatment with antiretroviral drugs when they are not acutely sick. To create a population that represents unwell (immuno-compromised) HIV-infected children initiating treatment for the first-time, we could have used only the first recorded weight and height.

Our method has some limitations. First, WHO provides a growth chart for a height up to 18 years, but for weight only up to 10 years. Hence, we used the CDC chart for weight in children above 10 years. However, there are differences between these charts. Where WHO charts describe the growth of healthy children under optimal environmental and health conditions measured in selected communities globally,\(^\text{12}\) CDC charts describe the growth of children during a certain historical time span. In 2015-2016, one in five US children between 11 and 19 was obese and this is reflected in the LMS parameters of the CDC tables.\(^\text{13}\) To mitigate this, we adjusted the CDC charts to be in line with WHO before calculating the functions for the adjustment factors. Second, the derived functions to adjust the growth charts are empirical and extrapolation to other populations (eg, tuberculosis or malaria patients, pre-term neonates) should be done with caution, taking any additional assumptions into account. A graphical comparison, as demonstrated in Figure 4, can be useful to investigate the agreement between a virtual population used for simulation and a target population. Only then one can decide to accept any deviations and proceed with the simulations or readjust the virtual population to better represent the target population.

Choosing a relevant population is an essential step in producing realistic simulation results. To allow for transparent and reproducible science, it is crucial that details about simulation results include a clear description of the population used, ideally making it publicly available. However, when using real patient data, this is not always possible due to data anonymity concerns. The workflow we provided here can be applied to generate in silico populations of interest that agree with an observed population, thus circumventing issues with data anonymity. This way, the population can be shared publicly, which supports full reproducibility of simulation results and reuse of the same population by future investigators.

ACKNOWLEDGEMENTS

We would like to thank the ARROW and CHAPAS-3 investigators for sharing their data for this work, as well as the staff for their contribution to the projects. Finally, we thank the children and their parents for their time and participation in these studies. This study is supported by PediCAP, which is part of the European and Developing Countries Clinical Trials Partnership (EDCTP) programme supported by the European Union (grant number RIA2017MC-2023).

COMPETING INTERESTS

R.W. and P.D. are supported by PediCAP, which is part of the EDCTP programme supported by the European Union (grant number RIA2017MC-2023). E.M.S. is supported by PanACEA, which is part of the EDCTP programme supported by the European Union (grant number TRIA2015-1102-PanACEA). A.S.W. and M.C. are supported by core support from the Medical Research Council UK to the MRC Clinical Trials Unit (MC_UU_12023/22) through a concordat with the Department for International Development. A.S.W. is also an National Institute for Health Research (NIHR) Senior Investigator. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR, the Department of Health or Public Health England (PHE). P.D. receives support from the National Research Foundation (NRF) of South Africa (NRF grant 109 056).

CONTRIBUTORS

R.E.W. participated in design, analysis and interpretation of the data and drafting the manuscript. E.M.S. and P.D. participated in conception, design, analysis and interpretation of the data and revising the manuscript. A.S.W. and M.N.C. participated in acquisition of data, analysis and interpretation of the data and revising the manuscript.

DATA AVAILABILITY STATEMENT

The virtual population and the R-script to generate this population are available in the Supporting Information.

ORCID

Roeland E. Wasmann https://orcid.org/0000-0002-5769-8150
Elin M. Svensson https://orcid.org/0000-0002-0993-6445
A. Sarah Walker https://orcid.org/0000-0002-0412-8509
Michelle N. Clements https://orcid.org/0000-0002-3442-3947
Paolo Denti https://orcid.org/0000-0001-7494-079X

REFERENCES


SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.