Population variability in GEC clearance

## Introduction

‘Reference ranges play an important role in clinical medicine, with values that lie outside the reference range viewed as an indication for further investigation and/or treatment.’ {Cole2009}

The analysis allows to compare the GEC of a single individual, or a group of people, to be compared with that of the simulated results for the reference population.

The precise pattern of change in body mass during adulthood varies from population to population.

“Data from cross-sectional studies (studies conducted on a population at one point in time) indicate a continual loss of height after young adulthood. This appears to be the net result of three factors. First, there is some actual shrinkage in height that begins during the fourth or fifth decade of life. Second, there has been a secular trend towards earlier maturation and larger body size since the early 1900s, although this trend appears to have slowed dramatically or stopped in many countries (Tanner, 1981). Third, among the elderly, there are fewer tall individuals and more short individuals than among young adults, i.e. there appears to be a differential mortality associated with body size (Lentner, 1984).

Results of a longitudinal study (subjects are followed over a period of time) conducted in Norway indicate a loss of height in both sexes of about 3 cm from young adulthood to age 70 years (Table 4.2). This is about one-half of the loss usually seen in cross-sectional studies, i.e. from combined secular and individual losses in height.” {Valentin2002}

For estimating interindividual differences in galactose clearance by the liver, knowledge of the distributions of input parameters is essential.

Evaluation of models for risk assessment applications includes considerations for **model purpose, model structure, mathematical representation, parameter estimation, computer implementation, and predictive capacity as well as sensitivity, variability and uncertainty analysis**. {EPA2006}

The mechnistic parameters, namely liver volume and hepatic blood flow were described with probability density functions (PDF) estimated from individual person and mean person data (LMS, GAMLSS) approach. Since each PDF depicts the frequency f occurence of all expected values for each parameter in the population, the effects of multiple sources of uncertainty and variability were accounted for in the estimated distribution of GEC in the population.

Necessary to account for the correlation between the parameters. Liver volume and and blood flow depending on age and gender were represented by (1) probability distribution functions that define both value of the parameters as well as frequency and (2) the correlations among the parameters to characterize the interrelationships among the parameters. In this study we considered the population characteristics sex and age. The distribution parameters are different for different ages and males and females.

## Estimating Intraspecies Variability - Human interindividual extrapolation

For estimating intraspecies variability, pupoulation distributions of partameters, particularly those relating to physiology and metabolizing enzymes (i.e. genetic polymorphisms), are specified in a Monte-Carlo approach, such that the model output corresponds to the distributions of values of interest in a population. Using the Monte Carlo approach, repeated computations based on inputs selected at random from statistical distributions for each input parameter are conducted to provide a statistical distribution of the output. Using high percentile (e.g. 95th) and 50th percentile, the intraspecies variability can be calculated.

The degree of human interindividual variance in key biochemical machinery produces differences in galactose clearance among humans. Effect of variation in metabolic enzymes on level on expression (dietary factors and lifestyle) and on level of activity (genetic polymorphisms)

## Sensitivity, variability and uncertainty analyses

### Sensitivity analysis

Sensitivity analysis provides a quantitative evaluation of how parameters in input functions influence the dose metric or outcome. Sensitivity analysis compares the magnitude in change in output for a defined change in each input parameter.. This process of single changes in one parameter while all others are held constant is called ‘local’ parameter sensitivity analysis.

(Problems of combined changes of parameters).

### Variability analysis

The focus of variability analysis is to evaluate the range of values that a parameter expected to be present in individuals may have in a population and the impact of that variability on the simulation prediction.

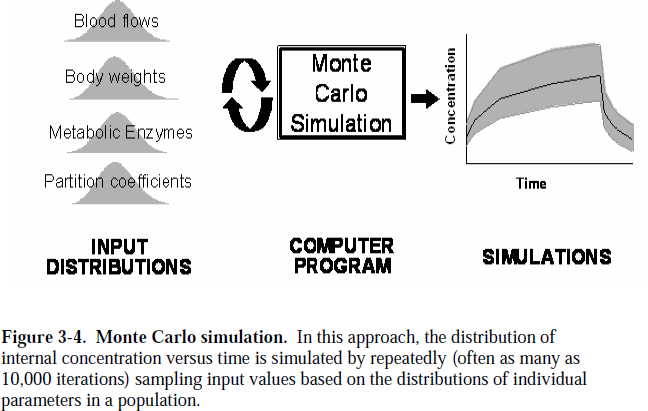
The magnitude of interindividual variability can be characterized using information such as the 95th percentile and 50th percentile in prediction. To derive this information, Monte Carlo simulations based on distributions of input parameters have frequently be used. (Lipscomb et al., 2003; Gentry et al., 2002; Haber et al., 2002; Lipscomb and Kedderis, 2002; Timchalk et al., 2002; Bogaards et al., 2001; El-Masri et al., 1999; Thomas et al., 1996a, b). The Monte Carlo method consists of repeated computation using inputs selected at random from statistical distributions for each parameter.

**Fit functions for physiological parameters**

Physiological parameters must be well characterized both in terms of measures of central tendency as well as variability. Necessity to provide reference values with associated variability in the population for physiological paramters based on individual-level data. {Johns2010} Especially important is therby the stratification based on age and gender. Age- and gender-related differences in physiological function can significantly affect the pharmacokinetics in various subpopulations of interest, namely young children and older adults.

Values of physiological parameters used in PBPK models are inherently variable and uncertain. The variability of these values represents, in part, the degree to which the parameters differ across a population. It is important to recognize and, to the extent possible, characterize this variability in order to better understand and characterize the variability in the overall model in which the parameters are used.

For many purposes, knowledge of the average value or the range of plausible values for model parameters is sufficient; however, for estimating interindividual differences in clearance, knowledge of the distributions of input parameters in the population is essential.



**Monte Carlo** simulation based on sampling from input distributions of individual parameters in the population. When conducting variability analysis, it is important to identify correlations in model parameters.

Probabilistic analysis was carried out using Monte Carlo simulation.

“A way to account for prior information on the model parameters invokes Monte Carlo simulation, requiring the following steps:

1. generation of sampling distributions for the assigned parameters, based on their mean and variance (or distribution), to produce a large number of “possible” values for these parameters. The correlation structure between the parameters must be taken into account to generate possible ‘parameter sets’.
2. solving the (kinetic) model for each set of these “possible” values to generate synthetic data, to which pseudorandom noise with characteristcs similar to the actual measurement noise should be added
3. identifying the model from the synthetic data and assessing the influence of variability of the assigned paraemters on the final variability of the desired parameter estimates.

“

{Vicini1999}

“Another approach which is increasingly used is the Monte Carlo Markov Chan (MCMC) model to quantify inter-individual variability.

**Predictive Modeling**

If a medical diagnostic is used for predicting aspects of disease (e.g, response to treatment, screening patients), patients desire the most accurate prediction possible. As long as complex models are properly validated, it may be improper to use a model that is built for interpretation rather than predictive performance.

### GAMLSS

“Generalized additive models for location, scale and shape (GAMLSS) are semi-paramtric regression type models. They are parametric, in that they require a parametric distribution assumption for the response variable, and ‘semi’ in the sense that the modelling of the parameters of the distribution, as functions of explanatory variables, may involve using nn-parametric smoothing functions.” {Stasinopoulos2007}

In [statistics](https://en.wikipedia.org/wiki/Statistics), the **generalized additive model location, scale and shape (GAMLSS)** is a class of [statistical model](https://en.wikipedia.org/wiki/Statistical_model) that provides extended capabilities compared to the simpler [generalized linear models](https://en.wikipedia.org/wiki/Generalized_linear_model) and [generalized additive models](https://en.wikipedia.org/wiki/Generalized_additive_model). These simpler models allow the typical values of a quantity being modelled to be related to whatever [explanatory variables](https://en.wikipedia.org/wiki/Independent_and_dependent_variables) are available. Here the "typical value" is more formally a [location parameter](https://en.wikipedia.org/wiki/Location_parameter), which only describes a limited aspect of the [probability distribution](https://en.wikipedia.org/wiki/Probability_distribution) of the [dependent variable](https://en.wikipedia.org/wiki/Independent_and_dependent_variables). The GAMLSS approach allows other [parameters](https://en.wikipedia.org/wiki/Statistical_parameter) of the distribution to be related to the explanatory variables; where these other parameters might be interpreted as [scale](https://en.wikipedia.org/wiki/Scale_parameter) and [shape parameters](https://en.wikipedia.org/wiki/Shape_parameter) of the distribution, although the approach is not limited to such parameters.

Cubic splines are fitted based on a method known as ‘backfitting’. In a generalized additive model, the weighted linear regression is simply replaced by a weighted backfitting algorithm. Details can be found in Chapter 6 of Hastie and Tibshirani {Hastie1990}. {Hastie1995}

### LMS

In GAMLSS the BCCG distribution is the **Box-Cox** transformation model used by Cole and Green (1992) (also known as the LMS method for centile estimation) {Stasinopoulos2007}

The principle of the LMS method is that the distribution of the outcome variable Y is defined by three age-varying parameters λ, μ, σ such that the transformation outcome

is a z-score with distribution close to N(0,1).

This distribution, called by GAMLSS the Box–Cox–Cole–Green (BCCG) distribution, has properties such that the distribution is symmetric, i.e. any skewness in Y is removed by suitable choice of the Box–Cox power λ, the location parameter μ is hence the median rather than the mean and the scale parameter σ is approximately the dimensionless CV or log standard deviation (SD). Kurtosis is assumed to be absent. {Cole2009}

GAMLSS is a generalization of the LMS method where Y has a specified frequency distribution

D(,, , ), the parameters representing the first four moments of the distribution

A wide variety of distributional forms are available, of which the normal distribution (called NO by GAMLSS)

is the simplest with just two parameters, location (mean ) and scale (SD ). Other distributions

have location, scale and one or two shape parameters (skewness and kurtosis ), including the

BCCG distribution (1) (where is equivalent to ) and the Box–Cox–power–exponential (BCPE)

distribution, which is an extension of the BCCG distribution to include kurtosis.

“The age trends for each moment are fitted using cubic spline curves, which are more flexible than polynomials or fractional polynomials for modelling complex nonlinear relationships.” {Cole2008}

### Centile estimation

Centile estimation includes methods for estimating the age related distribution of human growth.

The standard estimation of centile curves involves two continuous variables:

1. the response variable, that is, the variable we are interested in and for which we are trying to find the centile curves, e.g. weight, BMI, head circumference etc. and
2. the explanatory variable age.

The 100p centile of a random variable Y is the value y such that p(Y > y)=p, i.e. y= inv.cdf(p), where inv.cdf() the the inverse cumulative distribution function of of Y applied to p. Here we consider the conditional centile of Y given explanatory variable x (usually age). By varying x a 100p centile curve of y(x) against x is obtained. Centile curves can be obtained for different values of p. The World Health Organisation uses 100p=(3, 15, 50, 85, 97) in its charts and 100p=(1, 3, 5, 15, 25, 50, 75, 85, 95, 97, 99) in its tables.

This can be extended to more than one explanatory variable.

The methodology for creating growth centile references for individuals from a population comprises two different methods:

1. the **non parametric method of quantile regression** (Koenker, 2005; Koenker and Bassett, 1978, Koenker and Ng (2005), He and Ng (1999) and Ng and Maechler (2007))
2. the **parametric LMS** (i.e. Lambda, Mu and Sigma) method of Cole (1988), Cole and Green (1992) and its extensions for example see Wright and Royston (1997), van Buuren and Fredriks (2001), and Rigby and Stasinopoulos (2004, 2006).

Here we are dealing with the LMS method and its extensions. The LMS method, within GAMLSS, is equivalent of assuming the Box- Cox Cole and Green distribution (BCCG) for the response variable and fitting a smooth curves for μ, σ, and ν. The BCCG distribution is derived by assuming that Y, the response variable is a specific function of a random variable Z which has a (truncated) normal distribution. The BCCG distribution is suitable for positively or negatively skew data depending on the values of the parameter ν.

Rigby and Stasinopoulos (2004, 2006) extended the LMS method (which allows for skewness and but not for kurtosis in the data), by introducing the Box-Cox power exponential (BCPE) and the Box-Cox t (BCT) distributions respectively and called the resulting methods LMSP and LMST respectively. The BCPE assumes that the transformed random variable Z has a (truncated) exponential power distribution while BCT assumes that Z has a (truncated) t distribution.

More recently the function lms() is introduced for fitting centile curves in gamlss package.

‘To construct attained growth curves, the distributional properties of the anthropometric measurements must be studied and centile estimates derived within and across ages. Methods based on distributional assumptions have been used widely for their ability to produce z-scores and estimate extreme centiles more accurately’ {Borghi2006}

‘The World Health Organisation (WHO), following a reveiw of the methods available {Borghi2006}, chose to use GAMLSS for the analysis of its recently published growth standard {WHO. WHO Child Growth Standards: Methods and Development. WHO: Geneva, 2006.}’ {Cole2009}

For centile estimation the [WHO Multicentre Growth Reference Study Group](http://www.who.int/childgrowth/en) have recommended GAMLSS and the Box-Cox power exponential (BCPE) distributions[[4]](https://en.wikipedia.org/wiki/Generalized_additive_model_for_location,_scale_and_shape#cite_note-4) for the construction of the WHO Child Growth Standards.[[5]](https://en.wikipedia.org/wiki/Generalized_additive_model_for_location,_scale_and_shape#cite_note-5)[[6]](https://en.wikipedia.org/wiki/Generalized_additive_model_for_location,_scale_and_shape#cite_note-6)

**Calculation of the centile curves**

For the calculation of centiles 100α, for each alpha the centile curve yα against x, is obtained by finding the fitted values (mu^, sigma^, nu^, tau^) for each x (over a range of values of x) and substituting the values into



with t\_tau,alpha the 100 alpha centile of t\_tau. (exact formula in {Rigby2006}) {Stasinopoulos2007}

### Datasets & analysis

#### NHANES

Heinemann1999

## References

Borghi, E et al. "Construction of the World Health Organization child growth standards: selection of methods for attained growth curves." *Statistics in medicine* 25.2 (2006): 247-265.

Cole, TJ et al. "Age‐and size‐related reference ranges: A case study of spirometry through childhood and adulthood." *Statistics in medicine* 28.5 (2009): 880-898.

Cole, Timothy J, and Pamela J Green. "Smoothing reference centile curves: the LMS method and penalized likelihood." *Statistics in medicine* 11.10 (1992): 1305-1319.

Hastie, Trevor, and Robert Tibshirani. "Generalized additive models for medical research." *Statistical methods in medical research* 4.3 (1995): 187-196.

EPA2006, [Approaches for the Application of Physiologically Based ...](http://cfpub.epa.gov/ncea/cfm/recordisplay.cfm?deid=157668)

Hastie, Trevor J, and Robert J Tibshirani. *Generalized additive models*. CRC Press, 1990.

Johns, Douglas O et al. "Physiological Parameters and Databases for PBPK Modeling." *Quantitative Modeling in Toxicology* (2010): 107-134.

Rigby, Robert A, and D Mikis Stasinopoulos. "Using the Box-Cox t distribution in GAMLSS to model skewness and kurtosis." *Statistical Modelling* 6.3 (2006): 209-229.

Rigby, RA, DM Stasinopoulos, and V Voudouris. "Discussion: A comparison of GAMLSS with quantile regression." *Statistical Modelling* 13.4 (2013): 335-348.

Stasinopoulos, D Mikis, and Robert A Rigby. "Generalized additive models for location scale and shape (GAMLSS) in R." *Journal of Statistical Software* 23.7 (2007): 1-46.

Valentin, Jack. "Basic anatomical and physiological data for use in radiological protection: reference values: ICRP Publication 89." *Annals of the ICRP* 32.3 (2002): 1-277.

Vicini, Paolo et al. "Individual prior information in a physiological model of 2H8-toluene kinetics: an empirical Bayes estimation strategy." *Risk Analysis* 19.6 (1999): 1127-1134.