

# Package ‘SimKid’

May 7, 2026

**Title** Simulate Virtual Pediatrics using Anthropometric Growth Charts

**Version** 1.0.0

**Description** Simulate a virtual population of subjects that has demographic distributions (height, weight, and BMI) and correlations (height and weight), by sex and age, which mimic those reported in real-world anthropometric growth charts (CDC, WHO, or Fenton).

**License** GPL (>= 3)

**Encoding** UTF-8

**RoxygenNote** 7.3.2

**Depends** R (>= 2.10)

**LazyData** true

**Imports** dplyr (>= 1.1.4), ggplot2 (>= 3.5.1), magrittr (>= 2.0.3), msm (>= 1.7.1), randomizr (>= 1.0.0), rlang (>= 1.1.4), stats (>= 4.4.1), tidyr (>= 1.3.1), tmvtnorm (>= 1.6), withr (>= 3.0.0)

**Suggests** spelling, testthat (>= 3.0.0)

**Config/testthat/edition** 3

**BugReports** <https://github.com/Andy000000000000/SimKid/issues>

**URL** <https://github.com/Andy000000000000/SimKid>

**Language** en-US

**NeedsCompilation** no

**Author** Andrew Santulli [aut, cre],  
Enhanced Pharmacodynamics LLC [cph, fnd]

**Maintainer** Andrew Santulli <asantulli@epd-llc.com>

**Repository** CRAN

**Date/Publication** 2025-10-08 08:50:12 UTC

## Contents

calc_bmi_bsa	2
cdc0	3
cdc_ages2to20yr_correlations_by_sex_htcm_wtkg_allreplicates	4
cdc_ages2to20yr_correlations_by_sex_htcm_wtkg_summarized	5
fent0	6
get_seeds	7
grow_kid	7
htwt0	8
kid0	9
lms_calc	10
sim_kid	11
validate_kid	16
who0	17
<b>Index</b>	<b>18</b>

---

calc_bmi_bsa	<i>Calculate BMI and BSA</i>
--------------	------------------------------

---

## Description

BMI and BSA are calculated for a data frame that minimally has columns of HTCM and WTKG. Output columns match the definitions given by sim\_kid().

## Usage

```
calc_bmi_bsa(data = NULL)
```

## Arguments

data            A data frame with columns of HTCM and WTKG.

## Details

Equations and methods involved during the creation of virtual subjects.

## Value

A data frame with columns of BMI, BSA1, BSA2, and BSA3 added to data:

- BMI: Body mass index in kilograms per meter squared, rounded to 1 decimal place.
- BSA1: Body surface area in meters squared, rounded to 2 decimal places; calculated using the Mosteller equation.
- BSA2: Body surface area in meters squared, rounded to 2 decimal places; calculated using the Gehan and George equation.
- BSA3: Body surface area in meters squared, rounded to 2 decimal places; calculated using the DuBois equation.

### Calculation of body mass index

The equation for body mass index in kilograms per meter squared is  $BMI = WTKG / ((HTCM/100)^2)$ .

### Calculation of body surface area

The Mosteller equation (1) for body surface area in meters squared is  $BSA1 = \sqrt{WTKG * HTCM / 3600}$ .

The Gehan and George equation (2) for body surface area in meters squared is  $BSA2 = 0.0235 * (WTKG^{0.51456}) * (HTCM^{0.425})$ .

The DuBois equation (3) for body surface area in meters squared is  $BSA3 = 0.007184 * (WTKG^{0.425}) * (HTCM^{0.725})$ .

(1) Mosteller RD. Simplified calculation of body-surface area. *N Engl J Med.* 1987 Oct 22;317(17):1098. <doi: 10.1056/NEJM198710223171717.> PMID: 3657876. (2) Gehan EA, George SL. Estimation of human body surface area from height and weight. *Cancer Chemother Rep.* 1970 Aug;54(4):225-35. PMID: 5527019. (3) Du Bois D, Du Bois EF. A formula to estimate the approximate surface area if height and weight be known. 1916. *Nutrition.* 1989 Sep-Oct;5(5):303-11; discussion 312-3. PMID: 2520314.

### Examples

```
demo0 <- sim_kid()
demo <- calc_bmi_bsa(data = demo0)
```

---

cdc0

*U.S. Centers for Disease Control and Prevention (CDC) Growth Charts of Weight, Height, and BMI for Age*

---

### Description

Original CSV data files were manipulated into a more usable format.

### Usage

```
cdc0
```

### Format

cdc0:

A data frame with 1,398 rows and 16 columns:

**CHART** Growth chart label

**VAR** Demographic variable (WTKG is weight in kg, HTCM is height in cm, BMI is body mass index in  $kg/m^2$ )

**SEXF** Female sex indicator (0 is male; 1 is female)

**AGEGRP** Age group bucket in months

**L** Power in the Box-Cox transformation (calculation of VAR using age)

**M** Median (calculation of VAR using age)

**S** Generalized coefficient of variation (calculation of VAR using age)

**P3** 3rd percentile of the given VAR

- P5** 5th percentile of the given VAR
- P10** 10th percentile of the given VAR
- P25** 25th percentile of the given VAR
- P50** 50th percentile of the given VAR
- P75** 75th percentile of the given VAR
- P90** 90th percentile of the given VAR
- P95** 95th percentile of the given VAR
- P97** 97th percentile of the given VAR

### Source

<https://www.cdc.gov/growthcharts/cdc-data-files.htm>

---

cdc\_ages2to20yr\_correlations\_by\_sex\_htcm\_wtkg\_allreplicates

*Optimized Correlations between Z-scores of Weight and Height for Ages 2 to 20 years using U.S. Centers for Disease Control and Prevention (CDC) Growth Charts*

---

### Description

Methods of dataset creation:

1. Virtual subject age is created: 1000 males and 1000 females per each month of age ranging from 25 to 239 months.
2. Virtual subject height and weight are created using the CDC growth charts (i.e., LMS parameters) and BMI is calculated.
3. The height and weight distributions by age are constrained between the 0.1 and 99.9 percentiles (i.e., z-scores from -3 to 3).
4. The correlations between z-score of height and z-score of weight are optimized separately by sex and age using a 1-year age bucket (ex. correlation for ages 2-3 yr, 3-4 yr, etc.).
5. Percentiles of BMI (3rd, 10th, 25th, 50th, 75th, 90th, 97th) for the virtual population is compared to matching percentiles of observed BMI (i.e. the CDC growth chart of BMI vs age using the lower end of the 1-year age bucket) to calculate sum of squares.
6. R optimize function is used to minimize the sum of squares, providing the optimal correlation between z-scores of height and weight per sex and year of age.
7. This process is repeated 10x.

### Usage

cdc\_ages2to20yr\_correlations\_by\_sex\_htcm\_wtkg\_allreplicates

**Format**

cdc\_ages2to20yr\_correlations\_by\_sex\_htcm\_wtkg\_allreplicates:

A data frame with 360 rows and 4 columns:

**ITER** Iteration of the repeated optimization procedure

**SEXF** Female sex indicator (0 is male; 1 is female)

**AGEGRP** Age group in months

**HTWT\_COR** Optimized correlation between z-score of weight and z-score of height

**Source**

data-raw/cdc0.csv (data-raw/kid0.csv subset) and data-raw/htwt0.csv

---

cdc\_ages2to20yr\_correlations\_by\_sex\_htcm\_wtkg\_summarized

*Mean Correlations across the Ten Replicates of Optimization between Z-scores of Weight and Height for Ages 2 to 20 years using U.S. Centers for Disease Control and Prevention (CDC) Growth Charts*

---

**Description**

The mean correlations over the 10x replicates of optimization is calculated for use in the 'SimKid' package. In the 'SimKid' package the optimized correlations are validated by simulating virtual populations, calculating BMI statistics, and overlaying with the respective CDC BMI vs. age growth charts.

**Usage**

cdc\_ages2to20yr\_correlations\_by\_sex\_htcm\_wtkg\_summarized

**Format**

cdc\_ages2to20yr\_correlations\_by\_sex\_htcm\_wtkg\_summarized:

A data frame with 36 rows and 5 columns:

**SEXF** Female sex indicator (0 is male; 1 is female)

**AGEGRP** Age group in months

**NITER** Number of iterations (i.e., replicates) that contributes to the MEAN\_HTWT\_COR calculation

**NSUBJ\_AGEMO\_SEXF** Number of virtual subjects, in each replicate of the optimization procedure, per month of age and per sex that contributes to the MEAN\_HTWT\_COR calculation

**MEAN\_HTWT\_COR** Mean across the replicates of optimized correlation between z-score of weight and z-score of height

**Source**

data-raw/cdc\_ages2to20yr\_correlations\_by\_sex\_htcm\_wtkg\_allreplicates.csv

fent0

*Fenton Growth Charts of Weight for Age in Preterm Infants***Description**

Fenton growth charts for male and female weight vs. age were digitized up to 40 weeks (for full-term) from the literature (1-3) using 'PinPoint Digitizer' (4). Fitting of weight LMS parameters by age and sex was done in R using the optimize function (5) and the sum of squares statistic between digitized and predicted weight percentiles. (1) <https://ucalgary.ca/resource/preterm-growth-chart/preterm-growth-chart> (2) Fenton, T.R., Kim, J.H. A systematic review and meta-analysis to revise the Fenton growth chart for preterm infants. BMC Pediatr 13, 59 (2013). doi:10.1186/1471-2431-13-59 (3) Fenton, T.R., Nasser, R., Eliasziw, M. et al. Validating the weight gain of preterm infants between the reference growth curve of the fetus and the term infant. BMC Pediatr 13, 92 (2013). doi:10.1186/1471-2431-13-92 (4) <https://mhismail.github.io/PinPoint-Landing/> (5) <<https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/optimize>>

**Usage**

fent0

**Format**

fent0:

A data frame with 38 rows and 16 columns:

**CHART** Growth chart label**VAR** Demographic variable (WTKG is weight in kg)**SEXF** Female sex indicator (0 is male; 1 is female)**AGEGRP** Newborn age group bucket in weeks (PNA is postnatal age; GA is gestational age)**L** Power in the Box-Cox transformation (calculation of VAR using age)**M** Median (calculation of VAR using age)**S** Generalized coefficient of variation (calculation of VAR using age)**P3** 3rd percentile of the given VAR**P5** 5th percentile of the given VAR**P10** 10th percentile of the given VAR**P25** 25th percentile of the given VAR**P50** 50th percentile of the given VAR**P75** 75th percentile of the given VAR**P90** 90th percentile of the given VAR**P95** 95th percentile of the given VAR**P97** 97th percentile of the given VAR**Source**<https://ucalgary.ca/resource/preterm-growth-chart/preterm-growth-chart>

---

get_seeds	<i>Generate a reproducible vector of random seeds</i>
-----------	---

---

**Description**

Use the user-specified masterseed to generate a vector of randomly sampled seeds that is reproducible by calling the same masterseed at a future time.

**Usage**

```
get_seeds(masterseed = NULL, nseed = 1000)
```

**Arguments**

masterseed	An integer ranging from 1 to <code>.Machine\$integer.max</code> that sets an overall seed for the simulation to ensure reproducibility of the results. Defaults to no seed.
nseed	A positive integer that specifies the number of subjects to simulate. Defaults to 1000.

**Value**

A vector of numeric integers of length `nseed`

**Examples**

```
get_seeds(masterseed = 513, nseed = 10)
```

---

grow_kid	<i>Grow the simulated virtual subjects using anthropometric growth chart data</i>
----------	---

---

**Description**

Following creation of a virtual population using `sim_kid()`, each virtual subject grows from their baseline age. It is assumed that each virtual subject remains at the same respective percentiles of height- and weight-for-age-and-sex as they were at baseline. For example, if `sim_kid()` created a male 2 year old at the 25th percentile of height and the 30th percentile of weight, then if allowed to grow to 3 years old, this subject would be at the 25th percentile of height and 30th percentile of weight for 3 year old males according to the given anthropometric growth chart. Note that this function will not work for virtual preterm newborns created using the Fenton growth chart data. Note that this function will not allow virtual subjects ages 0 to 2 yr to grow past 2 years.

**Usage**

```
grow_kid(data = NULL, grow_time = 0, tstep = 1, age0isbirth = FALSE)
```

**Arguments**

<code>data</code>	A data frame created by <code>sim_kid()</code> .
<code>grow_time</code>	A non-negative numeric specifying the duration of time in months the virtual subjects are allowed to grow for. Will be rounded to the nearest month.
<code>tstep</code>	A positive numeric specifying the time step for growth in months. Default of 1. Will be rounded to the nearest month.
<code>age0isbirth</code>	Logical TRUE or FALSE matching the <code>sim_kid()</code> input option used. Default of FALSE.

**Value**

A data frame with columns matching those of `data` and the number of rows equal to  $nrow(data) * (1 + grow\_time / tstep) - nsubtract$ . Where `nsubtract` is the number of records with age greater than 240 months.

**Examples**

```
# growth for 1 year at monthly time step
demo0 <- sim_kid()
demo <- grow_kid(data = demo0, grow_time = 12)
```

---

htwt0	<i>U.S. Centers for Disease Control and Prevention (CDC) and World Health Organization (WHO) Growth Charts of Weight for Height</i>
-------	---

---

**Description**

Original CSV data files were manipulated into a more usable format.

**Usage**

```
htwt0
```

**Format**

htwt0:

A data frame with 474 rows and 16 columns:

**CHART** Growth chart label

**VAR** Demographic variable (HTWT denotes weight (kg) calculated using height (cm))

**SEXF** Female sex indicator (0 is male; 1 is female)

**HTCMGRP** Height group

**L** Power in the Box-Cox transformation (calculation of weight using height)

**M** Median (calculation of weight using height)

**S** Generalized coefficient of variation (calculation of weight using height)

**P3** 3rd percentile of weight

**P5** 5th percentile of weight

**P10** 10th percentile of weight  
**P25** 25th percentile of weight  
**P50** 50th percentile of weight  
**P75** 75th percentile of weight  
**P90** 90th percentile of weight  
**P95** 95th percentile of weight  
**P97** 97th percentile of weight

### Source

<https://www.cdc.gov/growthcharts/cdc-data-files.htm> and <https://www.cdc.gov/growthcharts/who-data-files.htm>

---

kid0	<i>Master Combined Dataset of Growth Charts of Weight, Height, and BMI for Age</i>
------	--

---

### Description

`cdc0`, `who0`, and `fent0` were combined into a single dataset.

### Usage

kid0

### Format

kid0:

A data frame with 1,536 rows and 16 columns:

**CHART** Growth chart label

**VAR** Demographic variable (WTKG is weight in kg, HTCM is height in cm, BMI is body mass index in kg/m<sup>2</sup>)

**SEXF** Female sex indicator (0 is male; 1 is female)

**AGEGRP** Age group

**L** Power in the Box-Cox transformation (calculation of VAR using age)

**M** Median (calculation of VAR using age)

**S** Generalized coefficient of variation (calculation of VAR using age)

**P3** 3rd percentile of the given VAR

**P5** 5th percentile of the given VAR

**P10** 10th percentile of the given VAR

**P25** 25th percentile of the given VAR

**P50** 50th percentile of the given VAR

**P75** 75th percentile of the given VAR

**P90** 90th percentile of the given VAR

**P95** 95th percentile of the given VAR

**P97** 97th percentile of the given VAR

**Source**

data-raw/cdc0.csv, data-raw/who0.csv, data-raw/fent0.csv

---

lms_calc	<i>Calculate the dependent variable (height or weight) using LMS parameters from anthropometric growth charts at a given Z score</i>
----------	--

---

**Description**

Calculate the dependent variable (height or weight) using LMS parameters from anthropometric growth charts at a given Z score

**Usage**

```
lms_calc(z = 0, l = NA, m = NA, s = NA)
```

**Arguments**

z	Numeric or numerical vector of Z score(s) associated with a given percentile of the dependent variable. Default of 0 (i.e., 50th percentile).
l	Numeric or numerical vector of L parameter(s) from an anthropometric growth chart.
m	Numeric or numerical vector of M parameter(s) from an anthropometric growth chart.
s	Numeric or numerical vector of S parameter(s) from an anthropometric growth chart.

**Value**

Numeric or vector of numeric dependent variable value(s). Value is calculated according to: if l (rounded to 6 decimal places) is equal to 0, then  $= l * \exp(s * z)$ ; otherwise  $= m * (1 + l * s * z)^{(1/l)}$ .

**Examples**

```
# calculate weight (kg) for a male 2 year old at the 50th percentile of
# weight for age and sex using CDC Growth Chart LMS parameters.
wtkg <- lms_calc(
  z = qnorm(50/100),
  l = -0.2165012,
  m = 12.74154,
  s = 0.1081660
)

# calculate weight (kg) for male 2 year old at the 25th and 50th percentiles
# of weight.
wtkg <- lms_calc(
  z = c(qnorm(25/100),
```

```

qnorm(50/100)),
l = -0.2165012,
m = 12.74154,
s = 0.1081660
)

# calculate weight (kg) for male at 50th percentile and female at 25th
# percentile of weight.
wtkg <- lms_calc(
  z = c(0, qnorm(0.25)),
  l = c(-0.6213197, -1.0244713),
  m = c(14.40263, 13.94108),
  s = c(0.1118745, 0.1194917)
)

```

---

sim_kid	<i>Create body size metrics for virtual subjects using anthropometric growth chart data and a distribution of age</i>
---------	---

---

## Description

Body size metrics (height, weight, BMI, and BSA) are created for a population of virtual subjects. The body size metrics reflect the anthropometric growth chart distribution(s) and correlations (ex. height vs weight) according to virtual subject age and sex. The assumed distribution of age (uniform or truncated normal) and probability that a given subject is female are specified by the user. For ages greater than 2 years, CDC growth charts are used. For ages birth to 2 years, either CDC (the default) or WHO growth charts can be used. Note that while CDC growth charts are used to prevent a jump discontinuity at 2 years, WHO growth charts are recommended for ages 0 to 2 years. For birth only (postnatal age of zero), Fenton growth charts for preterm can be used according to a distribution of gestational age. Note: when using Fenton growth charts, only body weight will be simulated.

## Usage

```

sim_kid(
  num = 1,
  agedistr = "unif",
  agemean = NULL,
  agesd = NULL,
  agemin = NULL,
  agemax = NULL,
  prob_female = 0.5,
  age0isbirth = FALSE,
  age0to2yr_growthchart = "CDC",
  age2to20yr_correlate_htwt = TRUE,
  htwt_percentile_min = NULL,
  htwt_percentile_max = NULL,
  masterseed = NULL
)

```

**Arguments**

num	A positive integer that specifies the number of subjects to simulate. Defaults to a single subject. For <code>agedistr = "nperage"</code> the number of subjects per growth chart age and sex bin.
agedistr	A string that specifies the distribution used to create virtual subject age. <ul style="list-style-type: none"> <li>• <code>unif</code> (the default): A uniform distribution of age with a range from <code>agemin</code> to <code>agemax</code>.</li> <li>• <code>norm</code>: A truncated normal distribution of age with a mean of <code>agemean</code>, a standard deviation of <code>agesd</code>, and a range from <code>agemin</code> to <code>agemax</code>.</li> <li>• <code>nperage</code>: An equal number of subjects per growth chart age and sex bin.</li> </ul>
agemean	A positive numeric greater than or equal to <code>agemin</code> and less than or equal to <code>agemax</code> that specifies the mean age when <code>agedistr = "norm"</code> is specified. <ul style="list-style-type: none"> <li>• Only used for <code>agedistr = "norm"</code>.</li> <li>• Units of postnatal age in months for <code>age0to2yr_growthchart = "CDC"</code> or <code>age0to2yr_growthchart = "WHO"</code>.</li> <li>• Units of gestational age in weeks for <code>age0to2yr_growthchart = "FENTON"</code>.</li> </ul>
agesd	A numeric greater than or equal to 0 that specifies the standard deviation of age when <code>agedistr = "norm"</code> is specified. <ul style="list-style-type: none"> <li>• Only used for <code>agedistr = "norm"</code>.</li> <li>• Units of postnatal age in months for <code>age0to2yr_growthchart = "CDC"</code> or <code>age0to2yr_growthchart = "WHO"</code>.</li> <li>• Units of gestational age in weeks for <code>age0to2yr_growthchart = "FENTON"</code>.</li> </ul>
agemin	A numeric that specifies the lower range of age. Defaults to the maximum allowable range if missing. <ul style="list-style-type: none"> <li>• Must be greater than or equal to 0 months of postnatal age for <code>age0to2yr_growthchart = "CDC"</code> or <code>age0to2yr_growthchart = "WHO"</code>.</li> <li>• Must be greater than or equal to 22 weeks of gestational age for <code>age0to2yr_growthchart = "FENTON"</code>.</li> <li>• Must be less than or equal to <code>agemax</code>.</li> </ul>
agemax	A numeric that specifies the upper range of age. Defaults to the maximum allowable range if missing. <ul style="list-style-type: none"> <li>• Must be less than 240 months of postnatal age for <code>age0to2yr_growthchart = "CDC"</code> or <code>age0to2yr_growthchart = "WHO"</code>.</li> <li>• Must be less than 41 weeks of gestational age for <code>age0to2yr_growthchart = "FENTON"</code>.</li> <li>• Must be greater than or equal to <code>agemin</code>.</li> </ul>
prob_female	A numeric value with an inclusive range of 0 to 1 that specifies the probability that a given virtual subject is female. Defaults to 0.5.
age0isbirth	A logical that specifies whether age equal to zero denotes birth. <ul style="list-style-type: none"> <li>• <code>TRUE</code>: Age of 0 is birth.</li> <li>• <code>FALSE</code> (the default): Age of 0 is ages from birth to less than one month.</li> <li>• Not applicable nor used for <code>age0to2yr_growthchart = "FENTON"</code>, for which postnatal age is always zero.</li> </ul>

**age0to2yr\_growthchart**

A string that specifies which anthropometric growth charts are used for ages less than or equal to 2 years old.

- "CDC" (the default): United States Centers for Disease Control and Prevention growth charts are used.
- "WHO": World Health Organization growth charts are used.
- "FENTON": Fenton growth charts for preterm newborns are used. This option is only available when simulating virtual subjects at birth (postnatal age = 0).

**age2to20yr\_correlate\_htwt**

A logical that specifies whether correlations, by sex and year of age, are implemented between simulated height and simulated weight for ages greater than or equal to 2 years old.

- TRUE (the default): Correlations are implemented between simulated height and simulated weight according to an identical internal-systems-data version of [cdc\\_ages2to20yr\\_correlations\\_by\\_sex\\_htcm\\_wtkg\\_summarized](#) located within the data folder.
- FALSE: Height and weight are simulated independently without any correlation(s). Note that this will likely result in unrealistic virtual subjects.

**htwt\_percentile\_min**

A numeric value that specifies the minimum allowed percentile of simulated height and weight, expressed as a decimal.

- Must be greater than or equal to 0.001.
- Must be less than `htwt_percentile_max` when `age2to20yr_correlate_htwt = TRUE`. Must be less than or equal to `htwt_percentile_max` when `age2to20yr_correlate_htwt = FALSE`.
- Defaults to 0.001 when `age0to2yr_growthchart = "CDC"` or `age0to2yr_growthchart = "WHO"`.
- Defaults to 0.01 when `age0to2yr_growthchart = "FENTON"` to avoid non-viable birth weights.

**htwt\_percentile\_max**

A numeric value that specifies the maximum allowed percentile of simulated height and weight, expressed as a decimal.

- Must be less than or equal to 0.999.
- Must be greater than `htwt_percentile_min` when `age2to20yr_correlate_htwt = TRUE`. Must be greater than or equal to `htwt_percentile_min` when `age2to20yr_correlate_htwt = FALSE`.
- Defaults to 0.999 when `age0to2yr_growthchart = "CDC"` or `age0to2yr_growthchart = "WHO"`.
- Defaults to 0.99 when `age0to2yr_growthchart = "FENTON"` to avoid non-viable birth weights.

**masterseed**

An integer ranging from 1 to `.Machine$integer.max` that sets an overall seed for the simulation to ensure reproducibility of the results. Defaults to no seed.

**Details**

Equations and methods involved during the creation of virtual subjects.

**Value**

A data frame with the number of rows equal to num (except for agedistr = "nperage") and columns of:

- ID: An integer ranging from 1 to num that serves as a virtual subject identifier.
- SEXF: An integer of value 0 for male or 1 for female.
- AGEMO: Postnatal age in months.
- AGE: Postnatal age in years.
- GAWK: Gestational age in weeks.
- WTKG: Body weight in kilograms, rounded to 2 decimal places.
- HTCM: Body height in centimeters, rounded to the nearest centimeter.
- BMI: Body mass index in kilograms per meter squared, rounded to 1 decimal place.
- BSA1: Body surface area in meters squared, rounded to 2 decimal places; calculated using the Mosteller equation.
- BSA2: Body surface area in meters squared, rounded to 2 decimal places; calculated using the Gehan and George equation.
- BSA3: Body surface area in meters squared, rounded to 2 decimal places; calculated using the DuBois equation.
- ZWTKG: The z-score of weight-for-height for ages 0 to 2 years, weight-for-age for ages greater than 2 years, and weight-for-gestational-age for newborns when using Fenton growth charts.
- ZHTCM: The z-score of height-for-age.
- PWTKG: The percentile of weight corresponding to the respective z-score.
- PHTCM: The percentile of height corresponding to the respective z-score.
- CHART: The anthropometric growth chart used. An error will be returned if the simulation fails.

**Calculation of simulated body height and weight**

The equation for simulated body height in cm (HTCM) or weight in kg (WTKG) is: if L (rounded to 6 decimal places) is equal to 0, then  $= M \cdot \exp(S \cdot Z)$ ; otherwise  $= M \cdot (1 + L \cdot S \cdot Z)^{(1/L)}$  (1).

Where L, M, and S are obtained, using the independent variables of sex (SEXF) and age bucket (AGEGRP), from identical internal-systems-data versions of the combined anthropometric growth chart datasets ([kid0](#) and [htwt0](#) located within the data folder). And where Z, the z-score respective to either the height or weight distribution, is randomly sampled for each virtual subject.

(1) <https://www.cdc.gov/growthcharts/cdc-data-files.htm>

**Simulation of z-scores for variability in height and weight**

For ages 0 to 2 years, correlations between height and weight are always implemented. This is done by simulating height using length-for-age growth charts (see [kid0](#) located within the data folder) and then simulating weight using weight-for-height growth charts (see [htwt0](#) located within the data folder). For ages greater than 2 years, correlations between height and weight were repeatedly optimized (see [cdc\\_ages2to20yr\\_correlations\\_by\\_sex\\_htcm\\_wtkg\\_allreplicates](#) located within the

data folder) and then summarized to the mean (see [cdc\\_ages2to20yr\\_correlations\\_by\\_sex\\_htcm\\_wtkg\\_summarized](#) located within the data folder). For ages greater than 2 years, the user can override the default behavior that includes correlations (as per an identical internal-systems-data version of [cdc\\_ages2to20yr\\_correlations\\_by\\_sex\\_h](#) between simulated height and weight using the `age2to20yr_correlate_htwt` input.

For ages 0 to 2 years and for ages greater than 2 years when simulating without correlations between height and weight: The z-scores are obtained independently for height and weight and for each virtual subject via random sampling from a truncated standard normal distribution using `msm::rtnorm()`.

For ages greater than 2 years when simulating with correlations between height and weight: The z-scores are obtained simultaneously for height and weight and for each virtual subject via random sampling from a truncated multivariate standard normal distribution using `tmvtnorm::rtmvnorm()`.

### Calculation of body mass index

The equation for body mass index in kilograms per meter squared is  $BMI = WTKG / ((HTCM/100)^2)$ .

### Calculation of body surface area

The Mosteller equation (1) for body surface area in meters squared is  $BSA1 = \sqrt{WTKG * HTCM / 3600}$ .

The Gehan and George equation (2) for body surface area in meters squared is  $BSA2 = 0.0235 * (WTKG^{0.51456}) * (HTCM^{0.42656})$ .

The DuBois equation (3) for body surface area in meters squared is  $BSA3 = 0.007184 * (WTKG^{0.425}) * (HTCM^{0.725})$ .

(1) Mosteller RD. Simplified calculation of body-surface area. *N Engl J Med.* 1987 Oct 22;317(17):1098.

<doi: 10.1056/NEJM198710223171717.> PMID: 3657876. (2) Gehan EA, George SL. Estimation of human body surface area from height and weight. *Cancer Chemother Rep.* 1970 Aug;54(4):225-35. PMID: 5527019. (3) Du Bois D, Du Bois EF. A formula to estimate the approximate surface area if height and weight be known. 1916. *Nutrition.* 1989 Sep-Oct;5(5):303-11; discussion 312-3. PMID: 2520314.

### Examples

```
# Simulate 1 subject with an age randomly sampled from a uniform
# distribution of ages ranging
# from 0 to 20 years using CDC growth charts.
df_kids <- sim_kid()

# Simulate 10 female 3 year old subjects with a seed set for reproducibility.
df_kids <- sim_kid(
  num = 10,
  agedistr = "norm", agemean = 36, agesd = 0,
  prob_female = 1, masterseed = 513
)

# Simulate 10 subjects (approximately 50% female) with ages ranging from
# 1 year to 2 years
# according to a uniform distribution of age using WHO growth charts.
df_kids <- sim_kid(
  num = 10,
  agedistr = "unif", agemin = 12, agemax = 24,
  age0to2yr_growthchart = "WHO"
```

```
)

# Simulate 1 subject per age bin and per sex using CDC growth charts
df_kids <- sim_kid(agedistr = "nperage")
```

---

validate_kid	<i>Validate the simulated virtual subjects to anthropometric growth chart data</i>
--------------	--

---

## Description

Following creation of a virtual population using `sim_kid()`, overlay scatter plots are used to validate that the virtual population is reflective of the respective anthropometric growth chart data.

## Usage

```
validate_kid(
  data = NULL,
  age0isbirth = FALSE,
  overlay_percentile = NA,
  alpha = 0.4
)
```

## Arguments

data	A data frame created by <code>sim_kid()</code> .
age0isbirth	Logical TRUE or FALSE matching the <code>sim_kid()</code> input option used. Default of FALSE.
overlay_percentile	NA (default) for no ribbon overlay of simulated percentiles. Or a numeric greater than 0 and less than 1 specifying the simulated percentile interval to overlay. For example, input of 0.90 would overlay the 5th and 95th percentiles of simulated data.
alpha	Numeric between 0 and 1 specifying the simulated data transparency in validation plots. Default of 0.4.

## Value

A list of 5 'ggplot2' plot objects.

## Examples

```
demo0 <- sim_kid() # single subject
validation_plots <- validate_kid(data = demo0)
```

---

who0 *World Health Organization (WHO) Growth Charts of Weight and Height for Age*

---

**Description**

Original CSV data files were manipulated into a more usable format.

**Usage**

who0

**Format**

who0:

A data frame with 100 rows and 16 columns:

**CHART** Growth chart label

**VAR** Demographic variable (WTKG is weight in kg, HTCM is height in cm)

**SEXF** Female sex indicator (0 is male; 1 is female)

**AGEGRP** Age group bucket in months

**L** Power in the Box-Cox transformation (calculation of VAR using age)

**M** Median (calculation of VAR using age)

**S** Generalized coefficient of variation (calculation of VAR using age)

**P3** 3rd percentile of the given VAR

**P5** 5th percentile of the given VAR

**P10** 10th percentile of the given VAR

**P25** 25th percentile of the given VAR

**P50** 50th percentile of the given VAR

**P75** 75th percentile of the given VAR

**P90** 90th percentile of the given VAR

**P95** 95th percentile of the given VAR

**P97** 97th percentile of the given VAR

**Source**

<https://www.cdc.gov/growthcharts/who-data-files.htm>

# Index

## \* datasets

- cdc0, [3](#)
- cdc\_ages2to20yr\_correlations\_by\_sex\_htcm\_wtkg\_allreplicates,  
[4](#)
- cdc\_ages2to20yr\_correlations\_by\_sex\_htcm\_wtkg\_summarized,  
[5](#)
- fent0, [6](#)
- htwt0, [8](#)
- kid0, [9](#)
- who0, [17](#)

  

- calc\_bmi\_bsa, [2](#)
- cdc0, [3](#), [9](#)
- cdc\_ages2to20yr\_correlations\_by\_sex\_htcm\_wtkg\_allreplicates,  
[4](#), [14](#)
- cdc\_ages2to20yr\_correlations\_by\_sex\_htcm\_wtkg\_summarized,  
[5](#), [13](#), [15](#)

  

- fent0, [6](#), [9](#)

  

- get\_seeds, [7](#)
- grow\_kid, [7](#)

  

- htwt0, [8](#), [14](#)

  

- kid0, [9](#), [14](#)

  

- lms\_calc, [10](#)

  

- sim\_kid, [11](#)

  

- validate\_kid, [16](#)

  

- who0, [9](#), [17](#)