

**Instructions on the use of the SAS program and R function to
calculate extended BMI z-scores and percentiles.**

Both the SAS and R program/function calculate sigma (scale parameter for the half-normal distribution, extended BMI percentile, extended BMIz, and the CDC LMS z-scores for weight, height, and BMI. Note that for BMIs $\leq 95^{\text{th}}$ percentile of the CDC growth charts, the extended values for BMI are equal to the LMS values. The extended values differ **only** for children who have a BMI $> 95^{\text{th}}$ percentile.

For both the SAS program and R function, please be careful with age - the units should be months and use the most accurate information available (e.g., 23.4928 months). For the SAS program, if you have completed months of age, as in NHANES data, add 0.5 to the value before running the program. If you have age in days, divide by 365.25/12 (=30.4375). In the R function, completed (integer) age in months is handled by specifying 'integer=TRUE' when calling the function (the default is FALSE). Additional information on age is at the CDC website for the SAS growth charts,

<https://www.cdc.gov/nccdphp/dnpao/growthcharts/resources/sas.htm>.

One thing to note is that in the extended BMIz is the inverse cumulative distribution function (CDF) of the extended BMI percentile. If the extended percentile is very close to 100, the qnorm function in R (SAS probit function) produces an infinite or missing value. This occurs if the extended BMI percentile is > 99.99999999999999 . It is likely these values represent data entry errors, such as a 48-month-old with a BMI > 39 . For these cases, extended BMIz is set to 8.21 in both R and SAS, a value that is slightly greater than the largest value that can be calculated.

The SAS program in the supplemental material is 'ahb_ext_bmiz.sas'.

This program is similar to the SAS program at <https://www.cdc.gov/nccdphp/dnpao/growthcharts/resources/sas.htm>, but includes sigma (scale parameter of half-normal distribution), extended BMI percentile, and extended BMIZ. The instructions for using this program are in the 'Instructions for SAS Users' on the CDC website.

Briefly, download the reference data file 'cdc_ref.sas7bdat' for the LMS parameters from the 4th paragraph at the CDC website. Assuming you've placed both the reference data file and 'ahb_ext_bmiz.sas' into 'c:/folder_path/CDC_data/', you can run the program with the following statements:

```
libname refdir 'c:/folder_path/CDC_data';  
data mydata; set name_of_your_dataset;  
%include 'c:/folder_path/CDC_data/ahb_ext_bmiz.sas';  
proc means data=_cdcddata; run;
```

Change the libname and %include statements to reflect the locations of the CDCref_d.sas7bdat and ahb_ext_bmiz.sas files. The names of the variables in your dataset must be sex (1 for boys / 2 for girls), agemos (months of age), weight (in kg), height (in cm), and BMI. Please read the information at <https://www.cdc.gov/nccdphp/dnpao/growthcharts/resources/sas.htm> about agemos.

The calculated variables will be merged with your original data in the _cdcddata output file. The new BMI metrics are labeled, but the ones that may be of most interest are BMIZ (LMS z-score), ext_bmiz (extended BMI z-score), and ext_bmip (extended BMI percentile).

The R function, 'ahb_ext_bmiz.R', calculates the sigma (scale parameter for the half-normal distribution, extended BMI percentile, extended BMIz, and the CDC LMS z-scores for weight, height, and BMI. Note that for BMIs \leq 95th percentile of the CDC growth charts, the extended values for BMI are equal to the LMS values. The extended values differ **only** for children who have a BMI $>$ 95th percentile.

The function assumes that you have a variable 'sex' (coded as 1=boys / 2=girls) and variables for age in months, weight (kg), height (cm), and BMI (kg/m²). Please be careful with age - the units should be months and use the most accurate information available (e.g., 23.4928 months. If age information is available only for the completed (integer) age in months (such as in data from NHANES), this is handled by specifying 'integer=TRUE' when calling the function (the default is FALSE). Additional information on age is at the CDC website for the SAS growth charts, <https://www.cdc.gov/nccdphp/dnpao/growthcharts/resources/sas.htm>.

One thing to note is that in the extended BMIz is the inverse cumulative distribution function (CDF) of the extended BMI percentile. If the extended percentile is very close to 100, the qnorm function in R produces an infinite value. This occurs only if the extended BMI percentile is $>$ 99.99999999999999. This occurs very infrequently, such as a 48-month-old with a BMI $>$ 39, and these BMIs likely represent data entry errors. For these cases, extended BMIz is set to 8.21, a value that is slightly greater than the largest value that can be calculated.

When calling the function, specify the names of the variables (sex, age in months, weight, and height) in your dataset without quotation marks. The function will load 5 packages: Hmisc, data.table, magrittr, dplyr, and labelled.

To use the R function,

(1) Download the CDC reference file (CDCref_d.csv) from the CDC website, <https://www.cdc.gov/nccdphp/dnpao/growthcharts/resources/sas.htm>. This CSV file is in the 4th paragraph. Save this reference file into a folder/directory, which will be specified with the 'ref_dir' argument.

(2) Source the R function. Assuming you've placed ahb_ext_bmiz.R in '~/.R/Growth_Charts/', the statement would be source('~/.R/Growth_Charts/ahb_ext_bmiz.R')

(3) For a dataset that contains completed age in months, such as NHANES, you can run the function with a statement such as,

```
newdata <- ahb_ext_bmiz(data,  
                        age=ridageex, integer=TRUE,  
                        wt=bmxwt, ht=bmxht, bmi=bmi,  
                        ref_dir= '~/R/Growth_Charts/Data/')
```

The `ref_dir` argument gives the location of the `CDCref_d.csv` file. The default for the `'integer'` argument is `FALSE`. When `integer=TRUE`, the function adds 0.5 to the completed number of months for each child.

For datasets that have more accurate information on age (e.g., 26.4367 months), one could use:

```
newdata <- ahb_ext_bmiz(data, age=age_mos,  
                        wt=weight, ht=height, bmi=bmi,  
                        ref_dir='~/R/Growth_Charts/Data/')
```

The output dataset will contain the z-scores and extended BMI z-score along with the other variables in your dataset (see next page).

Variables in output dataset (dataframe/data.table)

The variables in the output are labeled. Two ways to see these labels in R are `label(cdcdata)` and `str(cdcdata)`. The following table shows a description of the output variables from the R function and the SAS program:

Table 1

Variable	Description
waz, wp	LMS Weight-for-sex/age z-score and percentile
haz, hp	LMS Height-for-sex/age z-score and percentile
bmiz, bmip	LMS BMI-for-sex/age z-score and percentile
mod_waz	Modified z-scores (weight, height, and BMI) for identifying outliers (see the information in the SAS growth charts program website)
mod_haz	
mod_bmi	
bmip95	BMI expressed as a percentage of the 95 th percentile. A value of 120 often used as the cut point for severe obesity.
sigma	Scale parameter of the half-normal distribution
ext_bmip	Extended BMI percentile
ext_bmiz	Extended BMI z-score