Standardizing anthropometric measures in children and adolescents with new functions for egen

Suzanna Vidmar, John Carlin, and Kylie Hesketh
Clinical Epidemiology and Biostatistics Unit and
Centre for Community Child Health
Murdoch Childrens Research Institute and
University of Melbourne Department of Pediatrics
Royal Children's Hospital, Parkville, Victoria 3052, Australia

Tim Cole
Centre for Pediatric Epidemiology and Biostatistics
Institute of Child Health
London, UK

Abstract. A new function for `egen` has been developed to allow transformation of child anthropometric data to z-scores using the LMS method and the reference data available from the 1990 British Growth Reference and the 2000 US CDC Growth Reference. An additional function allows for children to be categorized according to body mass index (weight/height^2) using international cutoff points recommended by the Childhood Obesity Working Group of the International Obesity Taskforce.

Keywords: dm0004, z-scores, LMS, egen

1 Introduction

Comparing anthropometric data from children of different ages is complicated by the fact that children are still growing. (We do not expect the height of a 5-year-old to be the same as the height of a 10-year-old!) Clinicians and researchers are often interested in the question, "Is this child taller, shorter, or about average compared to other children of the same age?" Two sets of population-based reference data are now widely used to address this question: the 1990 British Growth Reference and the 2000 Centers for Disease Control and Prevention (CDC) Growth Reference in the United States. The British Growth Reference was developed using data from seventeen cross-sectional surveys conducted in the United Kingdom between 1978 and 1993 (Cole, Freeman, and Preece 1998). The CDC Growth Reference was developed using data from five national surveys conducted in the United States between 1963 and 1994 and from supplemental birth data collected from birth certificates and medical records. The CDC Growth Reference has recently been updated (Kuczmarski et al. 2000; *www.cdc.gov/growthcharts*). Both references tabulate values obtained by the LMS method, developed by Cole (1990) and Cole and Green (1992), that can be used to transform crude data to standard deviation z-scores. These are standardized to the reference population for the child's age and

sex (or for height/length and sex). The LMS transformation removes skewness and adjusts for physiological changes in anthropometric measures that occur with age. Stata provides a convenient environment in which to apply the age-specific (or height/length-specific) LMS values and generate z-scores for each child in a dataset, using the `egen` command.

Until recently, there was no standard definition of overweight and obesity for children. The Childhood Obesity Working Group of the International Obesity Taskforce recently recommended the use of body mass index (BMI) cutoff points to categorize children as normal weight, overweight, or obese based on age, gender, and BMI (Cole et al. 2000). These cutoff points were developed using large, nationally representative cross-sectional datasets from six countries: Brazil, Great Britain, Hong Kong, the Netherlands, Singapore, and the United States. They correspond to equivalent adult BMI cutoff points endorsed by the World Health Organization of BMI<25 kg/m^2 for normal weight, BMI 25–29.99 kg/m^2 for overweight, and BMI≥30 kg/m^2 for obesity. An additional function of `egen` categorizes children as normal weight, overweight, or obese according to these international cutoff points.

2   Syntax

`egen` $[type]$ *newvar* = `zanthro`(*varname*,*chart*,*version*) $[$`if` *exp*$]$ $[$`in` *range*$]$,
  <u>x</u>`var`(*varname*) `gender`(*varname*) `gencode`(<u>male</u>=*code*, <u>f</u>`emale`=*code*)
  $[$<u>age</u>`unit`(*unit*) <u>noc</u>`utoff`$]$

`egen` $[type]$ *newvar* = `zbmicat`(*varname*) $[$`if` *exp*$]$ $[$`in` *range*$]$, <u>x</u>`var`(*varname*)
  `gender`(*varname*) `gencode`(<u>male</u>=*code*, <u>f</u>`emale`=*code*) $[$<u>age</u>`unit`(*unit*)$]$

`by ...:` may not be used with either of these functions.

3   Functions

`zanthro`(*varname*,*chart*,*version*) calculates z-scores for anthropometric measures in children and adolescents according to US or UK reference growth charts. The three arguments are

  varname the variable name of the measure in your dataset for which z-scores are to be calculated (e.g., `height`, `weight`, or `bmi`).
  chart   a valid chart code (see the tables in section 5 for a list of valid codes).
  version either US or UK. Some charts are available for only one of these versions.

`zbmicat`(*varname*) categorizes children from 2 to 18 years as normal weight, overweight or obese, using BMI cutoffs recommended by the Childhood Obesity Working Group of the International Obesity Taskforce. BMI is in kg/m^2. This function generates a

variable with values 1, 2, and 3 labeled, respectively, as "Normal wt", "Overweight", and "Obese".

4   Options

xvar(*varname*) specifies the variable used (along with gender) as the basis for standardizing the measure of interest. This variable is usually age but can also be height or length when the measurement is weight; i.e., weight-for-age, weight-for-height, and weight-for-length are all available growth charts.

gender(*varname*) specifies the gender variable. It can be string or numeric. The codes for male and female must be specified by the gencode() option.

gencode(male=*code*, female=*code*) specifies the codes for male and female. The gender can be specified in either order, and the comma is optional. No quotes are required around the codes, even if the gender variable is a string.

ageunit(*unit*) gives the unit for the age variable and is only valid for measurement-for-age charts; i.e., omit this option when the chart code is wh or wl (see section 5). The *unit* can be day, month, or year. This option can be omitted if the *unit* is year, as this is the default.

nocutoff is optional. By default, any z-scores with absolute value greater than or equal to 5 (i.e., values that are 5 standard deviations or more away from the mean) are set to missing. This option forces calculation of all z-scores, allowing for extreme values in your dataset.

The decision to have a default cutoff at 5 standard deviations from the mean was made as a way of attempting to capture extreme data entry errors. Apart from this and setting to missing any z-scores where the measurement is a nonpositive number, these functions will not automatically detect data errors. Please check your data, as always!

5   Growth charts

For both CDC and British charts, each growth reference is summarized by three numbers, called L, M, and S, which represent the skewness, median, and coefficient of variation of the measurement as it changes with xvar(). Where xvar() is age, L, M, and S are tabulated at monthly intervals. For height and length, these parameters are tabulated at 1 cm intervals. Where a child's age/height/length occurs within these intervals, values of L, M, and S are obtained via a linear interpolation between successive values of xvar(). (The BMI cutoff points are tabulated at 6 monthly intervals. The cutoff point where a child's age occurs within these intervals is also obtained via linear interpolation.) Where xvar() is outside the permitted range, zanthro and zbmicat return a missing value.

In the British charts, the L value was fixed at 1 for height, head circumference, sitting height, and leg length, since these measures are normally distributed.

Table 1: 2000 CDC Growth Charts: Version US

Chart code	Description	Measurement unit	xvar() range
ha	height-for-age	cm	2 to 20 years
wa	weight-for-age	kg	0 to 20 years
ba	bmi-for-age	kg/m^2	2 to 20 years
hca	head circumference-for-age	cm	0 to 36 months
la	length-for-age	cm	0 to 36.5 months
wh	weight-for-height	kg	77 to 121.5 cm
wl	weight-for-length	kg	45 to 103.5 cm

Table 2: 1990 British Growth Charts: Version UK

Chart code	Description	Measurement unit	xvar() range
ha	height-for-age	cm	0 to 23 years
wa	weight-for-age	kg	0 to 23 years
ba	bmi-for-age	kg/m^2	0 to 23 years
hca	head circumference-for-age	cm	Males: 0 to 18 years
			Females: 0 to 17 years
sha	sitting height-for-age	cm	0 to 23 years
lla	leg length-for-age	cm	0 to 23 years

6   Examples

We illustrate with some data on a set of Australian children aged 5 to 10 years. Below, we show the relevant data for the first 5 children. The variable sex is coded male = 1, female = 2.

```
. list in 1/5, noobs
```

id	sex	ageyrs	bmi
1	1	6.6	15.5
2	1	8.4	20.1
3	2	5.7	16.8
4	1	7.5	19.9
5	2	9.7	26.8

In order to compare the BMI of the children in this sample, for instance, with respect to socioeconomic grouping or between two time points several years apart, the BMI values may be converted to standardized z-scores. The z-scores are created using the following command:

```
. egen zbmiuk = zanthro(bmi,ba,UK), xvar(ageyrs) gender(sex)
>              gencode(male=1, female=2)
```

For this analysis of Australian children, it might be equally appropriate to use the US reference standards:

```
. egen zbmius = zanthro(bmi,ba,US), xvar(ageyrs) gender(sex)
>              gencode(male=1, female=2)
```

To determine the proportion of children who are overweight or obese, each child can be categorized using the following command:

```
. egen bmicat = zbmicat(bmi), xvar(ageyrs) gender(sex)
>              gencode(male=1, female=2)
```

Here are the results for the first 5 children:

```
. list in 1/5, noobs
```

id	sex	ageyrs	bmi	zbmiuk	zbmius	bmicat
1	1	6.6	15.5	-.0172949	.0460011	Normal wt
2	1	8.4	20.1	1.921224	1.573253	Overweight
3	2	5.7	16.8	.8006477	.9584172	Normal wt
4	1	7.5	19.9	2.08305	1.729176	Overweight
5	2	9.7	26.8	2.803655	2.183658	Obese

An example where the age variable is coded in months and the outcome of interest is weight would be handled as follows:

```
. egen zwauk = zanthro(weight,wa,UK), xvar(agemth) gender(sex)
>              gencode(male=1, female=2) ageunit(month)
```

Using the US standards for height or length:

```
. egen zwhus = zanthro(weight,wh,US), xvar(height) gender(sex)
>              gencode(male=1, female=2)
. egen zwlus = zanthro(weight,wl,US), xvar(length) gender(sex)
>              gencode(male=1, female=2)
```

With the variable sex coded male = M, female = F:

```
. egen zbauk1 = zanthro(bmi,ba,UK), xvar(agemth) gender(sex)
>               gencode(male=M, female=F) ageunit(month)
. egen bcat = zbmicat(bmi), xvar(agemth) gender(sex)
>             gencode(male=M, female=F) ageunit(month)
```

Now, swapping the order in specifying the gender codes and omitting the comma:

```
. egen zbauk1 = zanthro(bmi,ba,UK), xvar(agemth) gender(sex)
>               gencode(female=F male=M) ageunit(month)
```

and using just the first letter in specifying each gender:

```
. egen bcat = zbmicat(bmi), xvar(agemth) gender(sex)
>            gencode(f=F m=M) ageunit(month)
```

NOTE: The code cannot be abbreviated. It must be typed into the `gencode()` option exactly as it appears in your dataset; e.g., `sex` is coded male = Male, female = Female:

```
. egen bcat = zbmicat(bmi), xvar(agemth) gender(sex)
>            gencode(f=Female m=Male) ageunit(month)
```

and forcing calculation of all z-scores:

```
. egen zbauk2 = zanthro(bmi,ba,UK), xvar(agemth) gender(sex)
>              gencode(f=Female m=Male) ageunit(month) nocutoff
```

7   References

Cole, T. J. 1990. The LMS method for constructing normalized growth standards. *European Journal of Clinical Nutrition* 44: 45–60.

Cole, T. J., M. C. Bellizzi, K. M. Flegal, and W. H. Dietz. 2000. Establishing a standard definition for child overweight and obesity worldwide: international survey. *British Medical Journal* 320(7244): 1240–1243.

Cole, T. J., J. V. Freeman, and M. A. Preece. 1998. British 1990 growth reference centiles for weight, height, body mass index and head circumference fitted by maximum penalized likelihood. *Statistics in Medicine* 17(4): 407–29.

Cole, T. J. and P. J. Green. 1992. Smoothing reference centile curves: The LMS method and penalized likelihood. *Statistics in Medicine* 11: 1305–19.

Kuczmarski, R. J., C. L. Ogden, L. M. Grummer-Strawn, K. M. Flegal, S. S. Guo, R. Wei, Z. Mei, L. R. Curtin, A. F. Roche, and C. L. Johnson. 2000. CDC growth charts: United States. *Advance Data* 314: 1–27.

About the Authors

Suzanna Vidmar is a Senior Research Assistant in the Clinical Epidemiology and Biostatistics Unit, Murdoch Childrens Research Institute and University of Melbourne Department of Pediatrics at the Royal Children's Hospital, Melbourne, Australia.

John Carlin is a Professor in the Departments of Pediatrics and Public Health at the University of Melbourne and Director of the Clinical Epidemiology and Biostatistics Unit.

Kylie Hesketh is a PhD student in the Centre for Community Child Health, Murdoch Childrens Research Institute and University of Melbourne Department of Pediatrics.

Tim Cole is Professor of Medical Statistics in the Centre for Pediatric Epidemiology and Biostatistics, Institute of Child Health, London, UK, and has published widely on the analysis of human growth data.