Correcting measurement error in BMI using non-linear equations

Daniel J. Dutton
Lindsay McLaren, PhD

Thursday, June 14, 2012
The problem

The easiest way to measure obesity in a large number of people at the same time is through self-reported body mass index (BMI).

People do not accurately report their height and weight\(^1\text{-}^7\).

This leads to mass suspicion about the value of self-reported BMI measures\(^8\). Which is a problem, since we have lots of self-reported data we would like to use.
The problem

The suspicion about self-reported BMI is not new (e.g., Rowland 1990). 

Adjusting self-reported values (height, weight, and BMI) so that they more closely resemble measured values has been attempted by many 10-13.

In Canada there is one correction equation method that has become quite popular 12,13. But it represents a “linear” correction and we know that the measurement error is “non-linear”.
“Linear” versus “Non-linear”

Basic idea: the measurement error in BMI increases as weight increases. It would be nice if a correction equation accounted for that.
“Linear” versus “Non-linear”

This language has to do with the nature of the adjustment as we move along the BMI scale.

“Linear” means we adjust heavy BMIs by the same amount per unit as we correct light BMIs.

“Non-Linear” means we adjust BMI by a changing amount per unit as we move along the BMI scale.
“Linear” versus “Non-Linear”
Red is a straight line, blue is a curve.

Linear versus Non-Linear Correction
A non-striking difference across reasonable BMI levels

Corrected BMI vs. Self-Reported BMI for Linear and Non-Linear Corrections
“Linear” versus “Non-linear”

Rates of change of the previous lines

Linear Versus Non-Linear Correction Equation

Rates of Change in the size of the correction

- Non-Linear Rate of Change
- Linear Rate of Change
The project - idea

We modeled the self-reported BMI of individual as a function of measured BMI and measurement error, which itself is influence by measured weight:

\[ BMI_{SR} = BMI_M \cdot e^{(W_M + \varepsilon)} \]

This equation can be used to model the “non-linear” relationship we know exists between self-reported and measured BMI, we just need to find the right parameters.
The project - data

Used CCHS data with measured and self-reported height and weight from 2005 and 2008 to see if a non-linear correction equation was worth using over the more common linear equations.

Data: 6294 useable observations (3086 males, 3208 females). Randomly split in half into an equation generating group and a test group.

Outliers! (People who had a very large measurement error in height or weight)
Outliers randomly split? No

- Generation Group
- Testing Group

Diagram showing the division of the dataset into generation and testing groups.
The project - results

After a some algebraic rearrangement of our original model, \( BMI_{SR} = BMI_M \cdot e^{(W_M + \varepsilon)} \), we were left with these estimated models:

Females:

\[
\ln(W_{SR}) = 0.697 + 0.811 \ln(W_M) + 0.001W_M
\]

Males:

\[
\ln(W_{SR}) = 0.710 + 0.788 \ln(W_M) + 0.002W_M
\]

(These are “complicated” models to solve for measured weight)
The project - results

So we can write a correction equation derived “simpler” model, yet different from the popular linear equation:

Females (non-linear):

\[ W_{Corrected} = e^{\left(\frac{\ln(W_{SR}) - .209}{.942}\right)} \]

⇒ Use with \( H_{SR} \) to make BMI Corrected

Females (linear – for reference)\(^{12}\):

\[ BMI_{Corrected} = -.12 + 1.05(BMI_{SR}) \]
The project – bottom line

I am proposing using a few more steps to get a corrected BMI measure. Is it worth it?

Depends on your research question.
BMI Distributions by measurement type - Females

0
0.02
0.04
0.06
0.08
0.1

0
10
20
30
40
50
60
70

Density

Measured BMI
Log Correction
Self-Reported BMI
Linear Correction
Log Correction (Simple)
The project – bottom line

I am proposing using a few more steps to get a corrected BMI measure. Is it worth it?

Depends on your research question.

If you are interested in whether or not an individual is above BMI 30 or not, then linear correction works fine. If you are interested in the entire BMI distribution then you may want to consider a non-linear correction like this one.
References