Relationships Among Fat mass, Fat-free Mass and Height in Adults: a New Method of Statistical Analysis Applied to NHANES Data

Richard F. Burton

School of Life Sciences, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, G12 8QQ, UK.

Number of text pages, plus bibliography, Nineteen, including 4 pages of bibliography

Number of tables: One

Number of figures, graphs and charts: One

Supplementary material: Two pages

Abbreviated title (running headline): **Fat mass, Fat-free Mass and Height in Adults:**

Dr. Richard F Burton,

Postal address: West Medical Building, University of Glasgow, Glasgow G12 8QQ, UK

Telephone: +44 (0) 141 334 9738

Fax: +44 (0) 141 330 5481

Email: richard.burton@glasgow.ac.uk

Grant sponsorship None

Relationships Among Fat mass, Fat-free Mass and Height in Adults: a New Method of Statistical Analysis Applied to NHANES Data

ABSTRACT

Objectives: The positive influence of fat mass (FM) on fat-free mass (FFM) has been quantified previously by various methods involving regression analysis of population data, but some are fundamentally flawed through neglect of the tendency of taller individuals to carry more fat. Differences in FFM due to differences in FM – and not directly related to differences in height – are expressed as ΔFFM/ΔFM, denoted K_F. The main aims were to find a sounder regression-based method of quantifying K_F and simultaneously of estimating mean BMI_0, the BMI of hypothetical fat-free individuals. Other, related, objectives were to check the linearity of FFM-FM relationships and to quantify the correlation between FM and height.

Methods: New statistical methods, explored and verified by Monte Carlo simulation, were applied to NHANES data. Regression of height^2 on FFM and FM produced estimates of mean K_F and indirectly of BMI_0. Both were then adjusted to allow for variability in K_F around its mean. Its standard deviation was estimated by a novel method.

Results: Relationships between FFM and FM were linear, not semi-logarithmic as is sometimes assumed. Mean K_F is similar in Mexican American men and women, but higher in men than women in non-Hispanic European and African Americans. Mean BMI_0 is higher in men than in women. FM correlates more strongly with height than has been found previously.

Conclusions: A more accurate way of quantifying mean BMI_0 and the dependence of FFM on FM is established that may be easily applied to new and existing population data.

KEY WORDS

Fat mass—fat-free mass relationship, fat-free mass, BMI, Forbes rule, NHANES

Relationships Among Fat mass, Fat-free Mass and Height in Adults: a New Method of Statistical Analysis Applied to NHANES Data

INTRODUCTION

Interrelationships among fat mass (FM), fat-free mass (FFM), total body mass (BM) and height have long been studied in relation to the body mass index, dieting and nutrition, to energy balance (e.g. Owen et al., 1986, 1987; Thomas et al., 2010), to health and to population differences. It is well known that body fat is associated with significant amounts of non-fat tissue involved in its mechanical and physiological support so that changes in FM are generally accompanied by changes in FFM leading to a correlation between them (e.g. Brožek, 1963; Burton, 2010, 2012;. Chaston et al., 2007; Dixon et al., 2007; Forbes, 1987; Heymsfield et al., 2011, 2014; Keys and Brožek, 1953; Webster et al., 1984). Quantifying normal relationships between FFM and FM may help one to judge whether loss of FFM becomes excessive and potentially hazardous during weight loss interventions (Chaston et al., 2007).

A particular concern here is the change or difference in FFM associated with, and due to, a given change or difference in FM. Average values of their ratio, ΔFFM/ΔFM (here denoted K_F), are estimated by regression analysis of data on FM, FFM and height for particular population samples. The same analysis simultaneously produces estimates of what has been denoted BMI_0 (Hruschka et al., 2013). This is the hypothetical mean BMI of the sampled individuals if they were fat-free. It does not correspond to malnutrition or starvation, but to extrapolation of FFM-FM relationships to a hypothetical fat-free condition. Mean BMI_0 is potentially useful in comparing BMI-fatness relationships in populations of differing body build worldwide.

As discussed more fully below, various related regression methods have been used in the past for estimating sample means of ΔFFM/ΔFM, or else its equivalent, the percentage contributions of fat and lean tissue to BM differences (Burton, 2010, 2012; Garrow and Webster, 1985; Gray and Bauer, 1991; Heitmann and Garby, 1998; Hruschka et al., 2013; Mingrone et al., 2001; Webster et al., 1984). The method of Burton (2010) involves multiple regression of FFM on FM and height^2. Useful as such studies have been, it has not been recognized that many of the methods, including the latter, are theoretically flawed. This is because there are two separate reasons why FFM tends to correlate with FM. Firstly, variations in FM tend to be accompanied by variations in FFM—the causal link, quantified as K_F, that is of particular interest here—and, secondly, for a given percentage body fat, taller people tend to have more of both FFM and FM. This paper establishes and illustrates a theoretically sounder regression method (called here for convenience the ‘new method’) than that of the ‘old method’ of Burton (2010).

The derivation and details of the relevant equations are given under ‘Materials and methods’. Development and validation of the necessary theory involved Monte Carlo modeling, the description of which is available in the Appendix. As some of the statistical procedures are novel, it is important to note that, tested on model data, they give correct estimates (or virtually correct estimates, given the inherent statistical variability) of the chosen, and therefore known, input parameters.

As implied above, the old method is inappropriate if taller individuals tend to carry more fat. Correlation coefficients for FM and height ($r_{FM.height}$) of only 0.0-0.1 have been published for adults (Heymsfield et al., 2007; Larsson et al., 2006; Owen et al., 1986, 1987), but such low values may not always pertain. A sounder way of estimating mean K_F and mean BMI_0 is therefore needed.

The old and the new methods both call for a linear relationship between FM and FFM, with ΔFFM/ΔFM independent of FM. However, Forbes (1987) and others since (see

Discussion) have regarded FFM as varying in proportion to log(FM). This view has been challenged (Burton, 2010), but remains an important issue to resolve.

The main points to be explored here are the estimation of mean K_F and mean BMI_0 using both new and published regression methods, the linearity of FFM-FM relationships, the correlation between FM and height and the possibility of differences between the sexes.

MATERIALS AND METHODS

Data

The data are from the National Health and Nutrition Examination Survey (NHANES) for 1999-2006 (Centers for Disease Control and Prevention). Individuals self-identified as non-Hispanic (NH) white, NH Black and Mexican American. Men and women were treated separately. Pregnant women were excluded. FM was estimated by whole-body DXA scans, and FFM calculated as body mass (BM) minus FM. The data used are those for ages 20-50 years. They are published with weightings used to improve the estimation of nationwide statistics representative of the US, but that is not the objective of the present study and the use of weightings is not helpful here (Hruschka et al., 2013; Korn and Graubard, 1991, 1995), especially as the six population subgroups are treated separately.

Statistics

Statistical analyses and modeling were carried out using Excel 2007 (Microsoft Corporation, Redmond, Washington, USA) and Datafit 8.0 (Oakdale Engineering, Oakdale, PA, USA). The units used are kg and m. Means and standard deviations (SDs) were calculated for height, BM, FFM, FM and % fat and also correlation coefficients for height and FM ($r_{FM.height}$). ‘Regression’ refers to ordinary least squares (OLS) regression except when otherwise specified.

In general, not all algebraically correct equations give true functional values when used in regression analysis, even with very large samples (Warton et al., 2006). Therefore, the validity of the methods for estimating K_F and BMI_0 was checked by Monte Carlo modeling. It is only a specific hypothetical example, called Model A, that is reported (see the Appendix and Table 1), but simpler versions were also used for exploring general points and checking statistical methods..

To test the linearity, FFM was plotted against FM for each of the six sex and ethnicity data sets together with the regression line. In addition, each set of data was divided into five groups according to FM (as equal as possible in number) with the means of FFM and FM then being plotted on the same graphs as the individual data. This allows a visual, rather than statistical, test of linearity.

Estimation of mean K_F and mean BMI_0

According to the algebraic model of Burton (2010, 2012) the FFM consists of two conceptually distinct components. One, denoted FFM_F, tends to vary in proportion to FM, with the ratio $\Delta FFM_F/\Delta FM$ being K_F, its values being independent of height. The other component, denoted FFM_H, is unaffected by FM but is taken as tending to vary in proportion to height^q such that

$$FFM_H \approx H.\text{height}^q \qquad (1)$$

The exact value of q makes little difference here (Burton 2012) so is taken as exactly two as in the BMI. The parameter H is then the same as BMI_0, the BMI of a hypothetical fat-free person (Hruschka et al. 2013). Accordingly,

$$FFM = FFM_F + FFM_H = K_F.FM + BMI_0.\text{height}^2 \qquad (2)$$

In the ‘old method’ this was used as a regression equation to estimate K_F and BMI_0 (= H). However, the results are invalid if FM is correlated with height because the terms (K_F.FM) and (BMI_0.height^2) would not then be independent (i.e. there would be collinearity). Results obtained by the old method are given here for comparison with those obtained by the new method, which is described below.

Parameter estimates obtained by regression are distinguished here by curly brackets, with accompanying subscripts referring to the relevant equation numbers —for example $\{K_F\}_4$ and $\{BMI_0\}_4$ in the case of Eq. (4) below.

The following equation is a rearrangement of Eq. (2) with an extra, numerical, term α:

$$\text{Height}^2 = [1/\text{BMI}_0].\text{FFM} - [K_F/\text{BMI}_0].\text{FM} + \alpha \quad (3)$$

It can be used as a regression equation, with $\{K_F\}_3$ calculated as the ratio of the two regression coefficients $[K_F/\text{BMI}_0]$ and $[1/\text{BMI}_0]$. However, the calculation does not produce the standard error (SE) of $\{K_F\}_3$. Therefore the following version of Eq. (4) is used instead, thus allowing the estimation of both mean K_F and its SE.

$$\text{Height}^2 = \{1/\text{BMI}_0\}_4.[\text{FFM} - \{K_F\}_4.\text{FM}] + \alpha \quad (4)$$

Formulated this way, Eq. (4) has the disadvantage that it needs to be fitted by an iterative method such as that of Datafit. Values of α are uninteresting and are not reported. For $\{1/\text{BMI}_0\}_4$ to give a valid estimate of-mean BMI_0, it was combined with the coefficient of multiple determination, R^2, thus:

$$\{\text{BMI}_0\}_5 = R^2/\{1/\text{BMI}_0\}_4 \quad (5)$$

(Recall that when any variable Y is regressed on another variable X and X is regressed on Y, the product of the two regression coefficients is the square of the correlation coefficient.)

Eqs. (4) and (5) may be understood as follows. For each individual, FFM_H equals $[\text{FFM} - K_F.\text{FM}]$. Therefore, the use of Eq. (4) is equivalent to regressing height^2 on FFM_H with the value of $\{K_F\}_4$ being such as to exactly maximize their correlation. If one knew the values of FFM_H, one might estimate BMI_0 by the reverse regression, of FFM_H on height^2 (*cf* Eq. (1) with $q = 2$ and $H = \text{BMI}_0$), but that would require knowledge of K_F. It is the fact that the product of the two regression coefficients, for height^2 on FFM_H and for FFM_H on height^2, would equal R^2 that allows BMI_0 to be estimated using Eq (5).

Modeling shows that, for large sample sizes, the estimates $\{K_F\}_4$ and $\{BMI_0\}_5$ are almost exactly correct (i.e. with small SEs), but only if K_F is the same for every individual. Variation of K_F around its mean results in underestimation of that mean and overestimation of mean BMI_0. However, approximate corrections can be achieved by the following novel method. There are two stages to this, of which the first is estimation of the SD of K_F, denoted $\{SD_{KF}\}_x$. (The 'x' refers, not to an equation, but to the procedure described next.)

In plots of FFM on FM (e.g. Figure 1), the scatter of values about the corresponding straight regression lines tends to increase with increasing FM and the degree of this heteroscedasticity increases with SD_{KF}. It is this property, explained in the next paragraph, that allows estimation of SD_{KF}. First the regression line is calculated. Then, with the squares of the residuals regressed on FM^2, the gradient of that regression line is $\{SD_{KF}\}_x^2$. (Confusingly, in both Figure 1 and in plots, not shown, of the squared residuals against FM^2, this heteroscedasticity is usually invisible to the eye, being obscured by a different statistical effect. This is because the span of the largest residuals for any particular narrow range of FM tends to increase with the number of individuals within that narrow grouping and the frequency distribution of FM is generally far from uniform. One's perception of the plotted data is dominated by the outliers.)

To understand the heteroscedasticity in simplistic terms, recall that FFM equals (FFM_H + K_F.FM), but treat FFM_H as constant. There is then a linear relationship between FFM and FM with data scatter due to variability in K_F. When FM is zero, the scatter in K_F.FM is zero. With increasing FM that scatter increases in proportion to FM – and so does the scatter of residuals around the fitted regression line of FFM on FM. This effect increases with SD_{KF}. Squaring of the residuals makes all the values positive.

The second stage produces the final estimates of mean K_F and mean BMI_0, denoted $\{K_F\}^*$ and $\{BMI_0\}^*$ respectively. Its purpose is to correct for the variability in K_F amongst individuals in the sample that makes $\{K_F\}_4$ slightly too low and usually makes$\{BMI_0\}_5$ slightly too high. The method involves modifying the original data set by adding to each individual FFM a quantity K_G.FM, where K_G has random values that are normally-distributed around a mean of zero. The chosen SD is arbitrary and not critical. Next the equivalent of $\{K_F\}_4$ – call it $\{K_F\}'_4$ – is estimated for the now-modified data set, If the process is repeated for other SDs of K_G, there is a negative straight-line relationship between $\{K_F\}'_4$ and the expression $[\{SD_{KF}\}_x^2 + (\text{SD of } K_G)^2]$. The value of $\{K_F\}^*$ is then found by linear extrapolation as the value of $\{K_F\}'_4$ when $[\{SD_{KF}\}_x^2 + (\text{SD of } K_G)^2]$ is zero. The validity of this extrapolation was confirmed by modeling. Finally $\{BMI_0\}^*$ is estimated as [(mean FFM) – $\{K_F\}^* \times$ (mean FM)]/(mean height^2).

To minimize the SEs of $\{K_F\}'_4$, the sample sizes (*n*) of 1,028-2,037 were increased by reduplication of the original data set by arbitrary factors of 7 to 16, the exact numbers being unimportant. Only then was each FFM increased by addition of K_G.FM.

RESULTS

[Table 1 here]

Table 1 shows summary data for each of the six real population samples and also for Model A. For the latter the true mean and SD of K_F were chosen to be 0.50 $\pm$ 0.10 with the true mean of BMI_0 being 15.9 kg/m^2 (see Appendix). Also shown are sample sizes (*n*) and the means and SDs of height, FFM, FM, BM and % fat, followed by the results of further statistical analysis. These include correlation coefficients for height and FM ($r_{FM.height}$).. The *estimated* true values of K_F and BMI_0 are denoted $\{K_F\}^*$ and $\{BMI_0\}^*$. Their SEs are unknown, but are unlikely to be much greater than those for $\{K_F\}_3$, and $\{BMI_0\}_4$ because the corrections to the latter are small.

The estimates $\{K_F\}^*$ and $\{BMI_0\}^*$ for Model A (i.e. 0.51 and 15.4 kg/m^2) match quite well the respective input values of 0.50 and 15.9 kg/m^2. The estimated SD of $\{K_F\}_4$, 0.11, is close to the input value of 0.10. Despite the large sample size used in modeling (n = 6111), the estimates $\{K_F\}^*$ and $\{BMI_0\}^*$ depend a little on the seeds chosen in generating the random numbers.

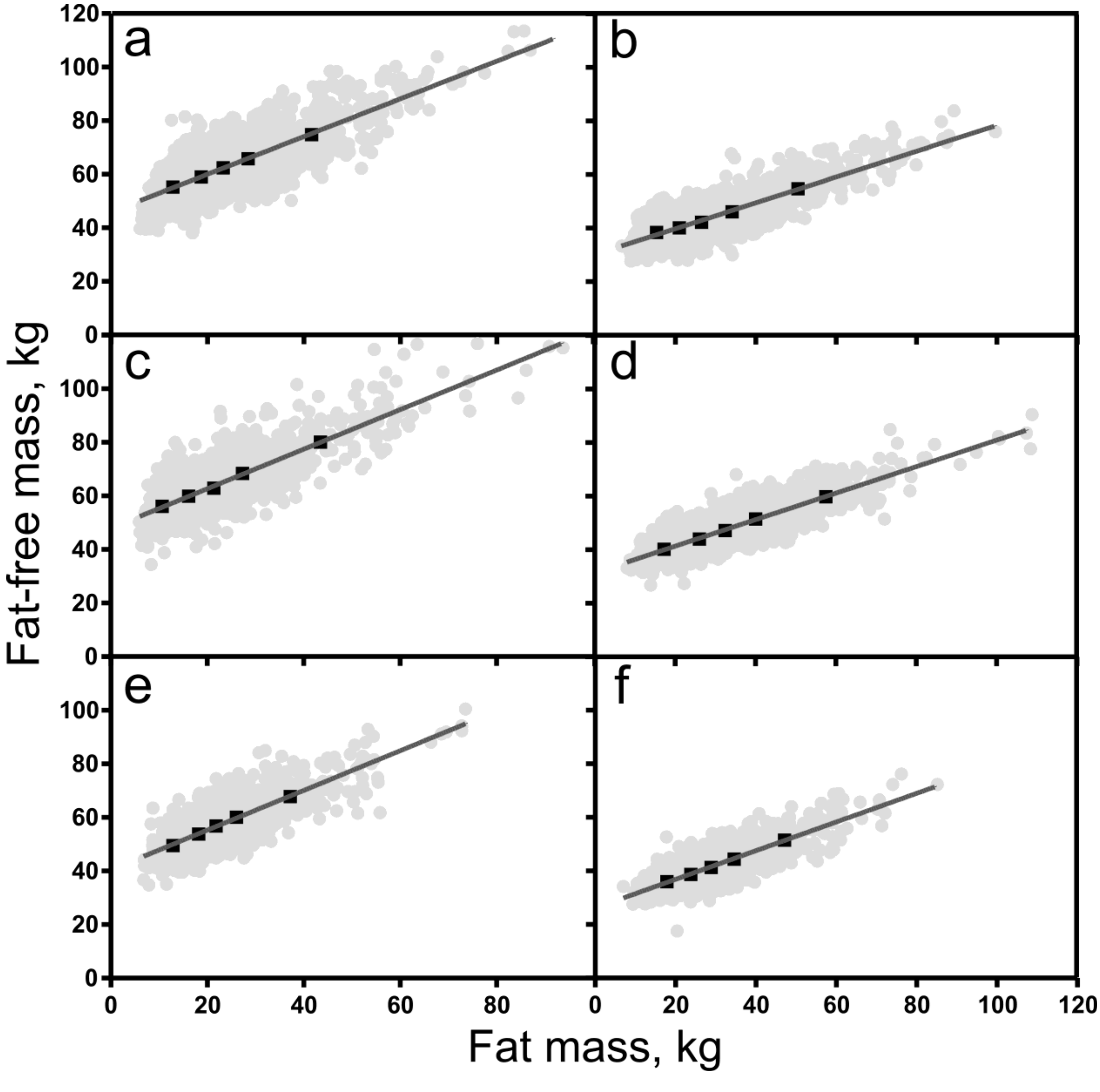

Figure 1

Figure 1 shows for all six population samples the relationship between FFM and FM. As well as the regression lines, the Figure shows mean values for data grouped according to FM. Visual inspection reveals only small and inconsistent deviations from linearity. This is not due to any methodological constraint. The lines accord with linear, not semi-logarithmic, relationships.

Estimates of mean K_F and mean BMI_0 obtained using regression equations (3) and (4) ($\{K_F\}_3$ and $\{BMI_0\}_4$) are respectively lower and higher than those obtained by the old method using regression Eq. (2), which are shown at the foot of Table 1. The adjusted new estimates ($\{K_F\}$* and $\{BMI_0\}$*) always fall between the old and (unadjusted) new estimates ($\{K_F\}_3$ and $\{BMI_0\}_4$). For the six real data sets the differences between $\{K_F\}_2$ and $\{K_F)$* and between $\{BMI_0\}_2$ and $\{BMI_0\}$* both correlate with $r_{FM.height}$, with correlation coefficients of 0.92 and –0.48 respectively. Regression of those differences on $r_{FM.height}$ indicates that, if $r_{FM.height}$ were zero, ($\{K_F\}_2$ – $\{K_F\}$*) would be about –0.01 and ($\{BMI_0\}_2$ – $\{BMI_0\}$*) would be about –0.3, with neither differing meaningfully from zero.

DISCUSSION

The main conclusion is that mean K_F and mean BMI_0 can be estimated for population samples by a new regression method, involving Eqs. (4) and (5) followed by estimation of, and correction for, the SD of K_F. The method is validated by Monte Carlo modeling. For the six samples, mean K_F was estimated after correction as 0.42-0.56 and corrected mean BMI_0 was estimated as 11.8-16.5 kg/m^2 (see $\{K_F\}$* and $\{BMI_0\}$* in Table 1). In the context of biological variation and measurement error, the corrections are quite small.

The regression method based on Eq. (2) (Burton 2010), described as the ‘old method’ is theoretically unsound and overestimates K_F and underestimates BMI_0. For reasons given in the Introduction, it only gives correct results when the correlation coefficient for FM and height, $r_{FM.height}$, is zero and it becomes less accurate as that correlation increases. This is demonstrable by Monte Carlo modeling and is illustrated by the present results. Values of $r_{FM.height}$ were found to be higher for the NHANES data than have previously been reported.

Fundamental to the modeling and statistical analysis is the assumption that FM and FFM are linearly related. Figure 1 shows that there is little if any systematic departure from linearity, but there is a contrary and influential view in the literature that the relationship is semi-logarithmic. These issues are discussed below, as well as previous research on FM-FFM relationships, differences between the sexes and the usefulness of BMI_0 as a concept.

The linearity of the FM-FFM relationships

In contrast to the present findings (Figure 1), Forbes (1987) described what he interpreted as a semi-logarithmic relationship for 164 females aged 14-50 years and similar in height (156-170 cm). In terms of kg his fitted curve has the following formula:

$$FFM = 23.9 \log_{10}(FM) + 14.2. \qquad (6)$$

However, the subjects included 12 with anorexia nervosa, which would presumably have reduced FFM_H as well as reducing FFM_F. The relationship for the other subjects, given their small number (152), is not convincingly non-linear (Gray and Bauer,1991) and there is much contrary evidence (Burton, 2010; Heitmann and Garby, 1998). Nevertheless, the semi-logarithmic nature of the relationship has been accepted by some later authors. Broyles et al. (2011) presented scattergrams of FFM plotted against ln(FM), separately for non-Hispanic European- and non-Hispanic African-American men and women, that seem to accord with it. However, the scatter in those four relationships and the limited ranges of FM and FFM make it unclear whether the relationships are best seen as linear or semi-logarithmic. Thomas et al. (2010) fitted an equation like Eq. (6) to NHANES data, but their plots of FFM/height^2 against FM/height^2 indicate linear rather than semi-logarithmic relationships. That is also true of their two graphs of best-fit fourth-order polynomial curves for FFM against FM. Webster et al. (1984) found straight-line relationships between BM/height^2 and FM/height^2. That most data indicate linear relationship implies the corollary that K_F (ΔFFM/ΔFM) is unaffected by FM.

The correlation between FM and height

As noted in the Introduction, previous evidence has suggested that $r_{FM.height}$ is typically 0-0.1. Moreover, it is 0.10 for data of Johnson (1996.) for 251 men. In contrast, Table 1 shows values of 0.12-0.33, which are all higher, as are the 95% CIs for the men. That they are lower for the women (0.12-0.14) than for the men (0.17-0.33) is unsurprising given that the FMs of women tend to be more variable. As can be shown by modeling and is illustrated by the real data, the errors associated with the old method, i.e. using Eq. (2), increase with $r_{FM.height}$ and are high enough to invalidate that method.

Other regression methods in the literature

A number of other regression equations have been published from which mean K_F may be estimated, including: BM/height^2 on FM/height^2 (Webster et al., 1984); FM/height^2 on BM/height^2 (Davies et al., 2001; Garrow and Webster, 1985; Gray and Fujioka, 1991); FFM on FM (Gray and Bauer, 1991); FFM on log(FM) (Mingrone et al., 2001). These, and others, have been discussed by Burton (2010). Of these authors only Gray and Bauer (1991) expressed their results as ratios equivalent to ΔFFM/ΔFM. Others did so in terms of percentage contributions of fat and lean tissue to BM differences. Davies et al. (2001) and Gray and Fujioka (1991) did not consider this issue.

Collectively, these studies have played a valuable role in establishing the correlation between FFM and FM, but not all produce reliable estimates of K_F. One reason is the smallness of some sample sizes (n = 24-198) and another is the unreliability of some methods of estimating FFM or FM. In addition, as noted in the Introduction, a fundamental problem is that FFM and FM are correlated for two distinct reasons.

Webster et al. (1984) applied a regression equation which, in terms of the present symbols, is:

$$[\text{BM/height}^2] = (1 + \{K_F\}_7) \times [\text{FM/height}^2] + \{\text{BMI}_0\}_7. \quad (7)$$

This may be derived from Eq. (2) by adding FM to both sides and then dividing both by height^2. In theory it produces the same estimates of K_F and BMI_0 as does Eq. (2) and Webster et al. (1984) noted that the slope of the regression line was little affected by 'correcting for height'. With the FMs being the means obtained by three different methods, mean $\{K_F\}_7$ for 104 women was 0.28. No meaning was attributed to the constant, $\{\text{BMI}_0\}_7$, which was 15.5 kg.m^2.

Garrow and Webster (1985) regressed FM/height^2 on BM/height^2, using what is essentially the 'reverse' of Eq. (7) and the same data set as used by Webster et al. (1984). Here K_F may be calculated from the correlation coefficient, r, and the regression coefficient, C, as (r^2/C – 1) = (0.955^2/0.713 – 1) = 0.28 as above. For a sample of 24 men it was (0.943^2/0.715 – 1) = 0.24.

Gray and Fujioka (1991) also regressed FM/height^2 on BM/height^2. Calculated as above from their equations, the mean values of K_F are 0.46 for 49 men and 0.25 for 140 women. Calculated from results of Davies et al. (2001) for 40 Chinese women—obtained by the same regression method—the mean K_F was 0.23. The correlation coefficients, $r_{\text{FM.height}}$, were not reported for either of these studies.

Heitmann and Garby (1998) used major axis regression of FM on FFM with the subjects of a given sex being of almost identical height (198 men with heights of 174.5-176.6 cm and 188 women with heights of 162.4-164.4 cm). Calculated from their results, $\{K_F\}$ is

0.47 for the men and 0.33 for the women (Burton, 2010). Because the heights were nearly constant, $r_{FM.height}$ must have been zero, making the estimates of K_F methodologically valid.

Regressing FFM on FM, Gray and Bauer (1991) obtained similar gradients ($\{K_F\}$), namely 0.57 for 29 men and 0.27 for 54 women for respective height ranges of 176-198 cm and 156-170 cm. The correlation coefficient $r_{FM.height}$ is unknown.

Burton (2012) regressed FFM/height^2 on FM/height^2 and also applied Eq. (2), using data for 59 men (Owen et al., 1987) and 31 women (Owen et al., 1986). The estimates of mean K_F from both methods were 0.50 for the men and 0.38 for the women. The correlation coefficients, $r_{FM.height}$, were not significant (i.e. 0.08 and 0.07 respectively). The values of $\{K_F\}^*$, calculated subsequently, are 0.44 for the men and 0.37 for the women. Regardless of significance levels, this fits the expectation that $\{K_F\}_2$ and $\{K_F\}^*$ should be similar when $r_{FM.height}$ is low.

Comparison of the sexes

The above studies of Gray and Fujioka (1991), Heitmann and Garby (1998), Gray and Bauer (1991) and Burton (2012), all indicate that mean K_F is higher in men (0.44-0.50) than in women (0.25-0.38), although, as noted, two of the pairs of estimates may have been affected by unknown correlations between FM and height. Regarding the present data (Table 1), $\{K_F\}^*$ is higher in men (0.54 and 0.56) than in women (0.42 and 0.43) for the non-Hispanic African and European Americans, but not for the Mexican Americans. Applying Eq. (2) to their New Zealand samples, Hruschka et al. (2013) found $\{K_F\}_2$ to be slightly higher in the women (0.43) than in the men (0.40). Both would be over-estimates if there were significant correlations between FM and height. In conclusion, there seems to be no clear general difference between the sexes in regard to K_F. Further evidence on possible differences between the sexes is discussed under the next heading, but it is again indecisive.

Evidence obtained from changes in body composition

Much of the evidence that FFM tends to change with FM has been obtained from alterations in body composition consequent on changes in nutrition, a topic usefully reviewed by Heymsfield et al. (2014). Chaston et al. (2007) reviewed changes in FFM during significant loss of BM, tabulating the effects of surgical interventions and, separately, the effects of dietary, behavioural and pharmaceutical interventions. Mean results are recorded as % loss of FFM (%FFML) which are readily translated to K_F, i.e. ΔFFM/ΔFM. This equals %FFML/(100 – %FFML). In regard to dietary and behavioral weight loss interventions, median values of K_F were 0.16 for ‘low calorie diets’, 0.31 for ‘very low calorie diets’ and 0.29 for ‘very low calorie diets with exercise’. Surgical interventions of various kinds produced medians of 0.21, 0.34 and 0.46. Numbers of subjects per group were naturally low (6-101) and there was considerable variability in K_F, with some values like those in Table 1 and some very different. The very low calorie diets produced more rapid loss of BM and it was suggested that a longer time would have been needed for attainment of a stable state. Indeed, the duration of such experiments generally may not always allow complete adjustment of FFM (Heymsfield et al., 2011).

Burton (2010) tabulated four mean values of K_F calculated from published changes in obese individuals on restricted diets that were excluded by Chaston et al. (2007) and for these the values were 0.21-0.56 (9-26 subjects per group). Also tabulated were results from three other studies on men involving overfeeding, these being 0.5, 0.54 and 0.65 (8-24 subjects per group) —results similar to $\{K_F\}^*$ in Table 1.

Regarding possible sex differences, Chaston et al. (2007) tabulated fewer results for men than for women. However, the authors noted that “there was a tendency for higher mean %FFML in cohorts of men (27 $\pm$ 7%) when compared with women (20 $\pm$ 8%, P = 0.08)”. Those means correspond to K_F values of 0.37 and 0.25 respectively. One may compare just the mean values of K_F that were paired for men and women, these being based on measured changes in obese individuals on restricted diets, but Chaston et al. (2007) reported few examples. Of these Burton (2010) tabulated the following paired values, for men and women respectively: 0.29 and 0.14 (Leenen et al., 1993); 0.39 and 0.31 (Hoie et al., 1993); 0.33 and 0.16 (Kockx et al., 1999). The statistical significance of the differences is not known.

Most clearly relating to the present study are the results of Heitmann and Garby (2002) based on long-term non-experimental changes; for 611 women mean K_F was 0.31 and for 636 men it was 0.50.

Relevance of BMI_0

The estimated mean values of BMI_0, the BMIs of hypothetical fat-free subjects, apply to population samples. They can only be estimated for individuals by assuming values of K_F. As expected, the estimates are all higher for men than for women, partly reflecting differences in muscularity (Hruschka et al., 2013). They are increased by muscle building (Hruschka, et al., 2013) and must be affected by other factors that influence BMI, such as relative leg length (Bogin and Beydoun, 2007; Norgan, 1994). They may be expected to vary from one population to another when these differ in typical body build, whether through nutrition, lifestyle or genetics. Mean values of BMI_0 should therefore be helpful in interpreting actual BMIs. Populations differing significantly in mean BMI_0 should be compared in regard to BMI by utilizing the ratio BMI/BMI_0 or its reciprocal, rather than differences ($BMI - BMI_0$). Thus the relationship between percentage body fat and BMI may be derived from Eq. (2) as follows:

$$\%\ \text{body fat} = 100/(1 + K_F) \times (1 - BMI_0/BMI) \qquad (8)$$

A concept related to BMI_0 is that of 'basal BMI', which is the (average) BMI of young adults in extremely poor rural households with few assets and where little excess body mass is accumulated (Hruschka et al., 2014). That has proved to be a useful concept in exploring ethnic differences and the influence of wealth through nutrition (Hruschka et al., 2014).

Conclusions

This paper is mainly about using regression analysis to estimate two conceptually distinct parameters relating to population samples, namely the mean of K_F, i.e. the value of ΔFFM/ΔFM when the influence of height is excluded, and the mean of BMI_0, the BMI of hypothetical fat-free individuals. Past regression methods of relating FFM, FM and height in these terms have mostly been theoretically invalid. The new method is sounder and accords with simple Monte Carlo modeling.

The new method is based on the assumption that the relationship between FFM and FM can be treated as linear, rather than as semi-logarithmic as is sometimes assumed. It is therefore important that new evidence for linearity has been obtained. Another finding relevant to the assessment of past regression results is that FM can be more closely correlated with height than has been reported previously.

The two parameters K_F and BMI_0 have been estimated for non-Hispanic European Americans, non-Hispanic African Americans and Mexican Americans, all aged 20-50 years. Mean K_F, as $\{K_F\}^*$, was estimated for these as 0.42 to 0.56. Much of the published evidence indicates that K_F tends to be higher in men than in women, but this is not a clear general rule.. In the case of the non-Hispanic European and African Americans K_F is indeed higher for the men than for the women, but that is not true of the Mexican Americans.

In their review, Heymsfield et al. (2014) have commented that the widely cited ‘Quarter FFM rule’ or ‘Forbes‘ Rule’, that weight loss composition consists of one fourth fat-free mass is “at best an approximation” and depends on a variety of circumstances ‘One fourth’ corresponds to K_F = 0.33, which matches some estimates, but is too low as a rule of thumb for the NHANES samples of Table 1. As expected, $\{BMI_0\}^*$ was found to be lower in the women (11.8-12.6 kg/m^2) than in the men (15.7-16.5 kg/m^2).

This new regression method can be easily applied to existing and future data sets, providing information on possible differences amongst populations worldwide.

ACKNOWLEDGMENTS

None.

There is no conflict of interest.

LITERATURE CITED

Bogin B, Beydoun N. 2007. The relationship of sitting height ratio to body mass index and fatness in the United States, 1988-1994. In: Singh SP, Gaur R, editors. Hum Ecol Special Issue. 15:1-8.

Broyles ST, Bouchard C, Bray GA, Greenway FL, Johnson WD, Newton RL, Ravussin E, Ryan DH, Smith SR, Katzmarzyk PT. 2011. Consistency of fat-mass—fat-free mass relationship across ethnicity and sex groups. Brit J Nutr 105:1272-1276.

Brožek J, Grande F, Anderson JT, Keys A. 1963. Densitometric analysis of body composition: revision of some quantitative assumptions. Ann NY Acad Sci 110:113-140.

Burton RF. 2010. The influence of fat mass on fat-free mass in healthy adults. Int J Body Comp Res 8:109-116.

Burton RF. 2012. Fat-free mass as a function of height and fat mass in healthy adults. Int J Body Comp Res 10:35-38.

Centers for Disease Control and Prevention (CDC). National Center for Health Statistics (NCHS). National Health and Nutrition Examination Survey Data. Hyattsville, MD: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention: http://www.cdc.gov/nchs/about/major/nhanes/dxx/dxa.htm.

Chaston TB, Dixon JB, O'Brien PE. 2007. Changes in fat-free mass during significant weight loss: a systematic review. Int J Obes 31:743-750.

Davies PSW, Lanham DA, Stead MA, Tsang K. 2001. Body mass index and the prediction of percentage body fat in Australian Chinese women, Ann Hum Biol 28:467-470.

Dixon JB, Strauss BJG, Laurie C, O'Brien PE. 2007. Changes in body composition with weight loss: obese subjects randomized to surgical and medical programs. Obesity15:1187–1198.

Forbes GB. 1987. Lean body mass—body fat interrelationships in humans. Nutr Rev 45:225-231.

Garrow JS, Webster J. 1985. Quetelet's index (W/H^2) as a measure of fatness. Int J Obes 9:142-153. .

Gray DS, Bauer M. 1991. The relationship between body fat mass and fat-free mass. J Am Coll Nutr 10:63-68.

Gray DS, Fujioka K. 1991. Use of relative weight and body mass index for the determination of adiposity. J Clin Epidemiol 44:545-550.

Heitmann BL, Garby L. 1998. Composition of body weight differences in subjects with the same body height Eur J Clin Nutr 52:619-620.

Heitmann BL, Garby L. 2002. Composition (lean and fat tissue) of weight changes in adult Danes. Am J Clin Nutr 75:840–847.

Heymsfield SB, Gallagher D, Mayer L, Beetsch J, Pietrobelli A. 2007. Scaling of human body composition to stature: new insights into body mass index. Am J Clin Nutr 86:82-91.

Heymsfield SB, Thomas D, Nguyen AM, Peng JZ, Martin C, Shen W, Strauss B, Bosy-Westphal A, Muller MJ. 2011. Voluntary weight loss: systematic review of early phase body composition changes.. Obes Rev 12:e348-e361.

Heymsfield SB, Gonzalez MCC, Shen W, Redman L, Thomas D. 2014. Weight loss composition is one fourth fat-free mass: a critical review and critique of the widely cited rule. Obes Rev 15:310-321.

Hoie LH, Bruusgaard D, Thom E. 1993. Reduction of body mass and change in body composition on a very low calorie diet. Int J Obes 17:17-20.

Hruschka DJ, Rush EC, Brewis AA. 2013. Population differences in the relationship between height, weight, and adiposity: an application of Burton's model. Am J Phys Anthropol 151:68-76.

Hruschka DJ, Hadley, Brewis A. 2014. Disentangling basal and accumulated body mass for cross-population comparisons. Am J Phys Anthropol 152:542-550.

Johnson RW. 1996. Fitting percentage of body fat to simple body measurements. J Stat Educ 4 [1]. (http://www.amstat.org/publications/jse/v4n1/datasets.johnson.html.)

Keys A, Brožek J. 1953. Body fat in adult man. Physiol Rev 33:245-325.

Kockx M, Leenen R, Seidell J, Princen HMG, Kooistra T. 1999. Relationship between visceral fat and PAI-1 in overweight men and women before and after weight loss. Thromb Haemost 82:1490-1496.

Korn EL, Graubard BL. 1991. Epidemiologic studies utilizing surveys: accounting for the sampling design. Am J Publ Health 81:1166-1173.

Korn EL, Graubard BL. 1995. Examples of differing weighted and unweighted estimates from a sample survey. Am Stat 49:291-295.

Larsson I, Henning B, Lindroos AK, Näslund I, Sjöström CD, Sjöström L. 2006. Optimized predictions of absolute and relative amounts of body fat from weight, height, other anthropometric predictors, and age. Am J Clin Nutr 83:252–259.

Leenen R, van der Kooy K, Droop A, Seidell JC, Deurenberg P, Weststrate JA, Hautvast JGAJ. 1993. Visceral fat loss measured by magnetic resonance imaging in relation to changes in serum lipid levels of obese men and women. Arterioscler Thromb 13:487-494.

Mingrone G, Marino S, DeGaetano A, Capristo E, Heymsfield SB, Gasbarrini G, Greco AV. 2001. Different limit to the body's ability of increasing fat-free mass. Metab 50:1004-1007.

Norgan NG. 1994. Interpretation of low body mass indices: Australian Aborigines. Am J Phys Anthropol 94:229-237.

Owen OE, Kavle E, Owen RS, Polansky M, Caprio S, Mozzoli MA, Kenrick ZV, Bushman MC, Boden G. 1986. A reappraisal of the caloric requirements in healthy women. Am J Clin Nutr 44:1-19.

Owen OE, Holup JL, D'Alessio DA, Craig ES, Polansky M, Smalley KJ, Kavle EC, Bushman MC, Owen LR, Mozzoli MA, Kenrick ZV, Boden GH. 1987. A reappraisal of the caloric requirements of men. Am J Clin Nutr 46:875-885.

Thomas D, Das SK, Levine JA, Martin CK, Mayer L, McDougall A, Strauss BJ, Heymsfield SB. 2010. New fat-free mass fat mass model for use in physiological energy balance equations. Nutr Metab (Lond) 7:39 (11 pages).

Warton DI, Wright IJ, Falster DS, Westoby M. 2006. Bivariate line-fitting methods for allometry. Biol Rev 81:259-291.

Webster JD, Hesp R, Garrow JS. 1984. The composition of excess weight in obese women estimated by body density, total body water and total body potassium. Hum Nutr Clin Nutr 38C:299-306.

APPENDIX

The Monte Carlo model

Described here are details of the modeling procedure, in which data are mostly generated from random numbers according to defined rules. The general statistical properties of the model do not depend on close realism, but Table 1 includes results obtained for a particular realistic model, denoted Model A. Simpler and less realistic versions of the model were also explored by varying the input parameters, though detailed results are not reported. Constructing Model A involved the following four steps.

Step 1. The first step was to prepare a plausible data set for height and FFM_H in a spreadsheet. First, a real set of values of height, total FFM and FM was taken, namely the data for non-Hispanic European-American men, this being the largest of the six sets. From each value of FFM was subtracted an amount equal to $\{K_F\}_4 \times$ FM, where $\{K_F\}_4$ is the mean value determined for those data using Eq. (4), i.e. 0.51. The original values of FM were then discarded. The number of 'individuals' was then tripled to 6111 by replicating all the 2037 pairs of height and FFM_H. The non-linear regression relationship between FFM_H and height was found to be:

$$FFM_H = 15.95.\text{height}^{1.99}.$$

With 1.99 rounded to 2.00, the equivalent relationship is:

$$FFM_H = 15.93.\text{height}^{2}.$$

Mean BMI_0 is 15.93 $\pm$ SD 1.77 kg/m^2.

Step 2. Alongside the paired values of height and FFM_H were tabulated two sets of normally distributed random numbers, $Rand_1$ and $Rand_2$, each with a mean of 1.0. Their role was to contribute scatter to the model data set as described below. Their SDs were chosen by trial and error to produce realistic results or else varied arbitrarily to explore properties of the model. The particular set of realistic results shown in Table 1 and called 'Model A' were obtained as follows. The total FM for each individual was calculated from the sum (S) of two quantities, F_1 and F_2. Of these, F_1, was calculated as [$0.3 \times \mathrm{FFM_H} \times Rand_1$] and is therefore correlated with $\mathrm{FFM_H}$ and so also with height. The second, F_2, calculated as [$10 \times Rand_2$], is uncorrelated with both $\mathrm{FFM_H}$ and height. Individual values of FM were calculated as [$0.036S^2$]. The squaring of S has no theoretical basis, but happens to produce a skewed frequency distribution for FM somewhat like that for the real data for non-Hispanic European-American men (the skewness for Model A being 0.9 and that for those real data being 1.4).

The sets of 'random' numbers, generated in Excel, are in fact pseudorandom and vary with the chosen 'seed'.

Step 3. The process so far produced individual values of height and FM, but not of FFM. Two parameters remain to be specified, namely the chosen mean value of K_F and its chosen standard deviation ($\mathrm{SD_{KF}}$). For Model A these are respectively 0.50 and 0.10. The individual values of K_F are the products of the chosen means of K_F and of random numbers from a third set that average 1.00. To complete the data set, total FFM was calculated for each individual as [$\mathrm{FFM_H} + K_F \times \mathrm{FM}$].

Step 4. The resulting artificial data set was analysed statistically like the real ones (see Materials and Methods), producing values of {K_F}, {$\mathrm{SD_{KF}}$} and {$\mathrm{BMI_0}$} to be compared with the known values of mean K_F, $\mathrm{SD_{KF}}$ and mean $\mathrm{BMI_0}$ as initially chosen. For Model A the latter are respectively 0.50, 0.10 and 15.9 kg/m^2.

TABLE 1. Results for the six population groups and for Model A: Non-Hispanic European Americans (NHEA), Non-Hispanic African Americans (NHAA) and Mexican Americans (MA).

	Men			Women			
	NHEA	NHAA	MA	NHEA	NHAA	MA	Model A
n	2037	1028	1168	1921	1055	1054	6111
Height, m	1.78	1.78	1.70	1.64	1.63	1.58	1.78
SD of height	0.069	0.071	0.072	0.062	0.064	0.063	0.069
BM, kg	88.4	89.2	80.8	73.8	83.1	72.8	86.9
SD of BM	19.6	22.9	17.0	20.1	22.7	17.2	20.2
FFM, kg	63.5	65.5	57.6	44.3	48.6	42.4	62.8
SD of FFM	10.0	11.7	8.8	7.8	8.8	7.3	10.6
FM, kg	24.9	23.7	23.2	29.5	34.6	30.4	24.1
SD of FM	10.8	12.5	9.2	13.2	14.8	10.8	11.0
% fat	27.2	25.2	27.8	38.4	40.1	40.8	26.5
SD of % fat	6.2	7.0	5.5	7.3	7.1	5.8	6.9
$r_{FM.height}$	0.17	0.21	0.33	0.12	0.12	0.14	0.20
95%CI	(0.13-0.21)	(0.15-0.27)	(0.28-0.38)	(0.08-0.16)	(0.06-0.18)	(0.08-0.20)	(0.18-0.22)
Results from regression using Eqs. (4) and (5)							
R^2	0.32	0.38	0.44	0.29	0.28	0.28	0.28
$\{K_F\}_4$	0.51	0.53	0.40	0.40	0.42	0.43	0.46
SE of $\{K_F\}_4$	0.02	0.03	0.03	0.01	0.02	0.02	0.02
$\{BMI_0\}_5$, kg/m^2	15.7	18.5	17.3	12.0	12.8	12.2	16.0

Estimates of SD of K_F, $\{SD_{KF}\}_x$, and final estimates of mean K_F, $\{K_F\}^*$, and mean BMI_0, $\{BMI_0\}^*$

$\{SD_{KF}\}_x$	0.09	0.09	0.10	0.05	0.05	0.07	0.11
95%CI	(0.07-0.11)	(0.05-0.12)	(0.07-0.12)	(0.03-0.06)	(0.03-0.06)	(0.05-0.08)	(0.10-0.12)
$\{K_F\}$*	0.54	0.56	0.46	0.42	0.43	0.46	0.51
$\{BMI_0\}$*, kg/m^2	15.7	16.5	16.2	11.9	12.6	11.8	15.9
Results by the 'old method', using Eq. (2)							
$\{K_F\}_2$	0.65	0.67	0.61	0.46	0.47	0.51	0.66
$\{BMI_0\}_2$, kg/m^2	14.9	15.7	15.0	11.4	12.0	10.8	14.7

n, number in sample; BM, body mass; FFM, fat-free mass; FM, fat mass; $r_{FM.height}$, correlation coefficient for FM and height ($P < 0.001$); R^2, coefficient of multiple determination for Eq. (4); $\{K_F\}$ and $\{BMI_0\}$, estimates of mean K_F and mean BMI_0, obtained using Eqs. (2), (4) and (5) as indicated by the subscripts; %CI, the 95% confidence limits of $r_{FM.height}$ and of $\{SD_{KF}\}_x$; $\{K_F\}$* and $\{BMI_0\}$*, the final estimates by the new method.

FIGURE LEGEND

Fig. 1. Relationships between fat-free mass (FFM) and fat mass (FM) for non-Hispanic European Americans (NHEA), non-Hispanic African Americans (NHAA) and Mexican Americans (MA). Grey circles: individual values. Black squares: mean values for data grouped in fat mass quintiles. The equations of the regression lines, and the correlation coefficients (*R*), are as follows.

a) NHEA men: FFM = 0.704.FM + 45.9. *R* = 0.76.

b) NHEA women: FFM = 0.483.FM + 30.1. *R* = 0.82.

c) NHAA men: FFM = 0.738.FM + 48.0. *R* = 0.79.

d) NHAA women: FFM = 0.495.FM + 31.4. *R* = 0.83.

e) MA men: FFM = 0.741.FM +40.4. *R* = 0.77.

f) MA women: FFM = 0.535.FM + 26.2. *R* = 0.80.