Statistical Methods in Medical Research

Analysing body composition as compositional data: An exploration of the relationship between body composition, body mass and bone strength

D Dumuid, JA Martín-Fernández, S Ellul, RS Kenett, M Wake, P Simm, L Baur, T Olds.

Abstract

Human body composition is made up of mutually exclusive and exhaustive parts (e.g., %truncal fat, %non-truncal fat and %fat-free mass) which are constrained to sum to the same total (100%). In statistical analyses, individual parts of body composition (e.g., %truncal fat or %fat-free mass) have traditionally been used as proxies for body composition, and have been linked with a range of health outcomes. But analysis of individual parts omits information about the other parts, which are intrinsically co-dependent because of the constant sum constraint of 100%. Further, body mass may be associated with health outcomes. We describe a statistical approach for body composition based on compositional data analysis. The body composition data are expressed as logratios to allow relative information about all the compositional parts to be explored simultaneously in relation to health outcomes. We describe a recent extension to the logratio approach to compositional data analysis which allows absolute information about the total of the compositional parts (body mass) to be considered alongside relative information about body composition. The statistical approach is illustrated by an example that explores the relationships between adults' body composition, body mass and bone strength.

Key words

Body composition; body mass; compositional data; logratio; multiplicative total

1. Introduction: Body composition is a composition!

Body composition and its associations with health are widely researched. Studies have linked body composition with a multitude of health-related outcomes, including

cardiovascular health,[1] inflammation,[2] bone strength[3] and mortality.[4,5] Yet previous studies have considered individual parts of body composition as a proxy for overall body composition. For example, the part of %body fat has regularly been used as an explanatory variable, without including the remaining part of %fat free mass. Some studies have divided %body fat into %truncal fat and %non-truncal fat, and have regressed these two components against health-related outcomes.[6,7] Other divisions of %body fat have included %android and %gynoid fat.[8] Usually, all of the body composition parts cannot be included in the same statistical model. There is good reason for this – including all parts (e.g., %truncal fat, %non-truncal fat, %fat-free mass) as explanatory variables in the same regression model induces perfect multi-collinearity and leads to singularity of the variance-covariance matrix and lack of convergence of estimators.

Body composition data, as their name implies, are compositional in form. All compositional datasets are conceptualised in *relative* terms (percentages), and made up of mutually exclusive and exhaustive parts that add up to a total of 100%. This constant sum constraint imposes built-in relationships – if one part is relatively decreased, one or more other parts must be relatively increased to maintain the total of 100%. For this reason, the set of constrained raw percentages are typically not used in traditional statistical models but they can be included as a set of logratios. The logratio approach was first described by John Aitchison in the 1980s,[9] and has since developed into a well-defined and established branch of statistics called compositional data analysis (CoDA). The application of CoDA within health research is in its infancy, and is yet to be explored in the context of body composition data. To our best knowledge, only one study (Dumuid et al, 2019) has used the CoDA approach to analyse body composition.[10] In Dumuid et al, body composition logratios were analysed as dependent variables, against 24-h activity composition logratios as independent variables. In this paper, we apply CoDA methods to body composition data analysed as explanatory variables. Section 2 reviews the logratio approach to compositional data. Section 3 considers including the absolute information of a body composition, Section 4 is a case study applied to real data. A concluding Section 5 discusses the results.

2. The logratio approach to body composition analysis

Body mass can be divided into several mutually exclusive and exhaustive parts, as directed by the research question. Consider a body composition vector $\boldsymbol{x} = [x_1,\ x_2, \ldots, x_D] \in \mathbb{R}^D$ with D positive parts (x_i) expressed in kg. When these parts are expressed as percentages of body mass, the resulting vector is subject to a constant sum constraint of 100%. Because of this, the sample space of compositional data is represented by a $(D-1)$-dimensional subset of the real space ($\boldsymbol{R}^D$), known as the D-part simplex ($\boldsymbol{S}^D$). When the body composition vector $\boldsymbol{x}$ is expressed as a set of logratios, it is transferred from the constrained simplex space to the unconstrained real space.[9] This is important because it means traditional statistical techniques not suitably applied in the constrained space can be applied in a meaningful way.

Several types of logratios have been defined. Two types of logratios are suitable for use in multiple regression models: isometric logratios (ILRs) and additive logratios (ALRs). Both sets of logratios (ILRs and ALRs) carry identical relative information and produce identical model-based predictions because they are related by a matrix product (proof in supplementary file). The ILRs have the advantage that when they are used in multiple regression models, individual beta coefficients can be interpreted. However, the practical meaningfulness of the interpretation of the ILR coefficients has been questioned.[11] The ALRs have the advantage that they are simple to construct, but beta coefficients for individual ALRs cannot be interpreted.[12] The decision to express the body composition vector $\boldsymbol{x}$ as ILRs or ALRs will depend on whether the research question requires interpretation of individual beta coefficients (ILRs) or model-based prediction (ILRs or ALRs).

2.1 Isometric logratios

The ILR representation of compositional data deals with vectors of coordinates in R^{D-1}, this matching the dimension of $\boldsymbol{S}^D$. To express compositional data as ILRs one can use the process called sequential binary partitioning (SBP).[13] In a SBP the body composition vector is sequentially divided into two groups of parts, until only one part remains in each group. That is, Step 1 divides the composition into two groups of

parts, where one group is coded +1, and the other group is coded –1. The subsequent step further divides one of the groups from Step 1 into two groups, again coding the parts in one group with +1, and the parts in the other group with –1. Uninvolved parts are coded 0. After $D-1$ steps, the SBP process is finalised. The coding values from each step of the SBP process contribute one row in the $(D-1)\times D$ sign matrix, *V*, as below

$$V=\begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_{D-1} \end{bmatrix} where\ f_k^T = \begin{bmatrix} e_{k1} \\ e_{k2} \\ \vdots \\ e_{kD} \end{bmatrix} for\ k=1,\dots,D-1\ and\ e_{kj}\ \epsilon\{-1,0,+1\}, j=1,\dots,D.$$

Each row of the sign matrix (*V*) defines the construction of one ILR. As there are $D-1$ rows in *V*, there will be $D-1$ ILRs. The first row corresponds to the first ILR. Subsequent rows of *V* define subsequent ILRs.

We define R_k, S_k & T_k as three mutually exclusive groups of parts $x_{j(j=1,\dots,D)}$ of vector $\boldsymbol{x}$ with the following elements in f_k

$$R_k = \{x_j |\ e_{kj} = +1\}, j=1,\dots,r_k\ ,$$

$$S_k = \{x_j |\ e_{kj} = -1\}, j=1,\dots,s_k\ ,$$

$$\text{and } T_k = \{x_j |\ e_{kj} = 0\}, j=1,\dots,t_k\ ,$$

where $|R_k| = r_k,\ |S_k| = s_k, |T_k| = t_k$, and $r_k + s_k + t_k = D$,for $k=1,\dots,D-1$.

The *k*-th step of the SBP divides the parts $R_k + S_k$ into two groups, one containing R_k parts (coded +1), the other containing S_k parts (coded –1). The geometric means of groups of parts are defined as

$$\tilde{x}_{R_k} = \left(\prod_{x_{j\epsilon Rk}} x_j\right)^{1/r_k} \text{ and } \tilde{x}_{S_k} = \left(\prod_{x_{j\epsilon Sk}} x_j\right)^{1/s_k} \quad . \qquad 1$$

The ILRs $z_{k\ (k=1,\dots,D-1)}$ are constructed as the logratio of $\tilde{x}_{R_k}$ to $\tilde{x}_{S_k}$, therefore the *k*-th ILR is defined as

$$z_k = \sqrt{\frac{r_k \cdot s_k}{r_k + s_k}} \ln\left(\frac{\tilde{x}_{R_k}}{\tilde{x}_{S_k}}\right), k=1,\dots,D-1\ , \qquad 2$$

where $\sqrt{\frac{r_k \cdot s_k}{r_k + s_k}}$ is a normalising constant, as described in Egozcue et al.[13] The ILR coordinates, $z_{k\ (k=1,\ldots,D-1)}$, take the same values regardless if the body composition vector $\boldsymbol{x}$ is expressed in its original units (kg) or in percentages, providing the relative information of body composition.

The ILRs can be thought of as a set of balances, where the groups of parts in the numerator are interpreted in relation to the groups of parts in the denominator.[13] There are many different ways that parts can be grouped together to create ILRs. The research question and theoretical knowledge must inform the SBP process and formation of ILRs so that estimates are directly interpretable. When the set of ILRs is used in statistical models, such as in multiple linear regression models, the beta coefficients corresponding to the ILRs reflect the estimated change in outcome when the groups (i.e., the geometric mean) of parts in the numerator are increased at the expense of the groups (i.e., the geometric mean) of parts in the denominator, whereas the other logratios remain constant.[14] Relative changes in geometric means of groups of parts may not always be relevant to the research question. For example, the research may aim to estimate the effect of exchanging one part for one other part while keeping the remaining parts constant. Such an aim cannot be addressed by interpreting individual ILR coefficients. Instead, model-based predictions from either ILRs or ALRs are required.[15]

2.2 Additive logratios

The ALR are a set of logratios $\boldsymbol{a}$, defined as

$$\boldsymbol{a} = [a_1, \ldots, a_{D-1}] = \ alr(\boldsymbol{x}) = \ \left[\ln\left(\frac{x_1}{x_D}\right), \ln\left(\frac{x_2}{x_D}\right), \ldots, \ln\left(\frac{x_{D-1}}{x_D}\right)\right]. \qquad 3$$

It is important to note that the ALR are asymmetric and not isometric, because the same part (x_D) is in the denominator of all logratios. The ALRs do not retain distances and angles in the Aitchison geometry, which imposes some limitations on their use in statistical applications.[16] In addition, the individual regression parameters $\beta_1, \beta_2, \ldots, \beta_{D-1}$ cannot directly be interpreted as the estimated change in the response variable when one ALR is changed independently of the remaining ALRs. This is because it is impossible to vary one ALR independent of the remaining ALRs, since

the part x_D is present in all ALRs. Thus, direct interpretation of regression parameters is not meaningful.

2.3 Prediction with logratio regression models

When using logratio regression models for prediction, the focus is not on interpreting individual regression beta coefficients. As relevant to the research question, predicted values of a health outcome can be compared across different body compositions. For example, predicted bone strength for high %body fat and low %fat-free mass could be compared to predicted bone strength for low %body fat and high %fat-free mass. In addition, the set of logratio regression parameters can be used to calculate the estimated outcome for incrementally differing body compositions across an empirical range (i.e., a predictive grid). Interpolation between the predicted points on the grid creates a response surface. This is a particularly useful approach for 3-part predictive compositions, as the relationship between body composition and an outcome can easily be visualised in a 2-dimensional ternary plot. Predictive models are assessed by predictive error, i.e. comparing measured outcomes to predicted outcomes in the sample data, via cross validation or with separate training and validation datasets.

3. Body composition data analysis: CoDA with a total

The logratio approach allows inclusion of all the body parts in the same model. This is important to comprehensively examine the relationship between body composition and health outcomes. However, this approach assumes that body composition is purely relative data. In CoDA, absolute values (mass in kg) are irrelevant, as the information is relative and the results are the same when the data are expressed in percentages. However, it is likely that both relative and absolute information is important to health. In particular, body mass varies in the population, and this variation may explain some variance in health outcomes. To enable the inclusion of both relative and absolute information in the same statistical model, an extension of the logratio approach is required.

Recently, a statistical framework for compositional data that extends CoDA in order to also consider the absolute information, tCoDA, was introduced by Pawlowsky Glahn et al.[17] In Coenders et al,[14] the authors showed that including an additional term

for the sum of the log-transformed vector together with the set of logratios to regression models enabled both relative and absolute information to be considered, independently of each other. In the case where the body composition vector is considered explanatory, the term for absolute information (t_i) should be computed as the normalised *multiplicative total* of all D absolute values, defined as

$$t_i = \frac{1}{\sqrt{D}}\ln(x_{i1}x_{i2}\cdots x_{iD}) = \frac{1}{\sqrt{D}}(\ln(x_{i1}) + \ln(x_{i2}) + \cdots + \ln(x_{iD})). \quad 5$$

where $x_{i1}, x_{i2}, \cdots, x_{iD}$ are the absolute values (kg) of each body part. In the case of body composition, the total term t_i can be considered as a proxy for the total absolute mass of a participant.

3.1 Using tCoDA with isometric logratios

The tCoDA ILR multiple linear regression model for *n* observations $(\boldsymbol{x}_i, y_i), i = 1,2,\ldots,n$, is

$$y_i = \beta_0 + \sum_{j=1}^{D-1}\beta_j z_{ij} + \beta_D t_i + \varepsilon_i,$$

with intercept β_0, regression parameters $\beta_1, \beta_2, \ldots, \beta_D$ and error $\varepsilon \sim N(0, \sigma^2)$, independently,[14] and z_{ij} and t_i are defined as earlier.

The regression parameters for the individual ILR coordinates can be interpreted as estimated change in the outcome *y* when the respective balance of one group of parts to another is changed, while the remaining balances and multiplicative total is kept constant. For the remaining balances to be kept constant, the parts in the numerator all increase in the same proportion, while the parts in the denominator all decrease in the same proportion. To ensure multiplicative total is kept constant, the increases in the numerator are compensated for by the decreases in the denominator. The regression parameter β_D indicates the estimated change in the outcome when body composition is kept constant but the multiplicative total changes.

3.2 Using CoDA with a total (tCoDA) for prediction

As in CoDA, the tCoDA multiple linear regression model will predict identical values for y_i regardless of whether ILRs or ALRs are used to represent the relative information of the body composition.[14]

For example, considering an ALR tCoDA regression model, for a 3-part predictive body composition [truncal fat; non truncal fat; fat free mass] of [6 kg; 18 kg; 36 kg], the estimated outcome *y* would be calculated as

$$\dot{y}_{[6;18;36]} = \beta_0 + \beta_1 \cdot \ln\left(\frac{6}{36}\right) + \beta_2 \cdot \ln\left(\frac{18}{36}\right) + \beta_3 \cdot \frac{1}{\sqrt{3}} \ln(6 \cdot 18 \cdot 36).$$

By using t_i as defined above, the absolute information (multiplicative total of parts, i.e., kg) can be varied by the application of a constant, k, while the relative information (the set of logratios) will remain unchanged when k is applied:

$$\dot{y}_{[6;18;36]} = \beta_0 + \beta_1 \cdot \ln\left(\frac{6 \cdot k'}{36 \cdot k'}\right) + \beta_2 \cdot \ln\left(\frac{18 \cdot k'}{36 \cdot k'}\right) + \beta_3 \cdot \frac{\mathbf{1}}{\sqrt{\mathbf{3}}} \ln(6 \cdot 18 \cdot 36 \cdot k),$$

where $k' = (k)^{1/3}$. This enables the association between health and the absolute information to be examined, while keeping the relative information (the body composition) constant. The regression parameter corresponding to the total term β_3 in the above example can be interpreted as the independent effect of t_i. An example of when this situation may occur empirically is during certain kinds of sports training (e.g., distance swimmers) where both high fat and high muscle are needed. When the athlete stops swimming, weight loss could occur without necessarily altering body composition.

Conversely, the association between health and changes to the relative information (the body composition) can be examined, while keeping the multiplicative total amount constant. An example of this situation may be observed during ageing, when body mass may remain static while body composition changes (bone and muscle are replaced by fat).

While the two situations above may be of interest, it is possible that they are uncommon in real life. It may be expected that as people's body mass varies, so does their body composition. Expected changes in body composition for increasing body mass could be determined from a separate regression model, and used as new data for predictions from the tCoDA model. Higher order terms (e.g., quadratics) and

interactions between body composition and multiplicative total could be explored and included in the prediction models.

The extended logratio approach for tCoDA is applied to real data in Section 4.

4. Example: The associations between adults' body composition and bone strength

The example uses data from the Child Health CheckPoint,[18] a cross-sectional study nested within the Longitudinal Study of Australian Children (LSAC).[19] Parents and guardians of participating children were invited to be involved in the study. Table 1 shows that there were much higher participation rates among female parents/guardians (mothers) than male (fathers). Data were collected from adult participants between February 2015 and March 2016. Ethical approval for the study was received from The Royal Children's Hospital (Melbourne) Human Research Ethics Committee (HREC33225) and the Australian Institute of Family Studies Ethics Committee (AIFS14-26). Written informed consent was obtained.

4.1 Measures

This study includes bone strength and body composition data from adult participants who were able to attend one of CheckPoint's mobile 'Comprehensive' Assessment Centres set up in seven major Australian cities. At these centres, bone strength was assessed by two scans of the dominant lower leg (Stratec XCT 2000L pQCT scanner, Pforzheim, Germany). The images were processed by Stratec XCT 2000 software (Version 6.20C), which calculated various bone strength measures. In this example, we quantify bone strength with the polar stress-strain index, a composite score describing bone strength derived from multiple measures incorporating the density of the bone and its geometrical parameters (area, cortical thickness etc). Details regarding the bone strength measure can be found elsewhere.[20] For analysis, polar stress-strain index was standardised to a z-score, with higher values indicating better bone strength. Body mass (in kg) and body composition were measured using four-limb bioimpedance analysis scales (InBody230, Biospace, Seoul, Korea).[21] Body composition was operationalized as three body parts; %truncal fat, %non-truncal fat and %fat-free mass. Covariates of age and sex were derived from a questionnaire. Composite family-level socioeconomic position was represented by a z-score

previously constructed by LSAC and comprising information about occupation, household income and education.[22] A total of 1090 participants provided complete data and were included in the subsequent analyses. Characteristics of included participants are shown in Table 1.

Table 1: Participant characteristics

Measures	Summary Statistics (n = 1090)
Covariates	Arithmetic means (SD)
Age (y)[a]	44.5 (5.0)
Socioeconomic position *z*-score	0.3 (1.0)
SEIFA (socioeconomic score)	1032 (58)
Sex	n (%)
Female	939 (86)
Male	151 (14)
Body Composition (kg)	Arithmetic means (SD)
Truncal fat	12.7 (5.3)
Non-truncal fat	11.9 (5.1)
Fat-free mass	49.2 (9.1)
Body Composition (%)	Arithmetic means (SD)
Truncal fat	16.8 (4.7)
Non-truncal fat	15.7 (4.4)
Fat-free mass	67.5 (8.9)
Body Composition (%)	Closed geometric means[b]
Truncal fat	16.4
Non-truncal fat	15.4
Fat-free mass	68.2
	Arithmetic means (SD)
Body mass (kg)	73.8 (14.8)

Total term[c]	5.0 (0.5)
Bone strength: Polar Stress-Strain Index (mm^3)	2565 (584)

SEIFA: Socio-Economic Indexes for Areas; Index of Relative Social Disadvantage Score, capturing area-based inequality. The score is standardised to have a national (Australian) mean of 1000 and SD of 100. Higher score indicates higher socioeconomic status. [a]Age range = 28.8 to 68.8 years. [b]Compositional descriptive statistics include the centre (geometric means of parts, normalised to sum to 100%. Descriptive statistics for dispersion are not univariate such as standard deviations, they are multivariate and are presented in a variation matrix (Table 2). [c]Multiplicative total term defined as $t = \frac{1}{\sqrt{3}}\ln(truncal\ fat\ (kg) \cdot\ non-truncal\ fat(kg) \cdot fat-free\ mass(kg))$

4.2 Descriptive exploration of body composition and body mass

The body composition data are plotted in a ternary diagram below (Figure 1). The centre (compositional sample mean) is represented by a black dot, surrounded by 50, 90 and 95% Gaussian prediction ellipses. The distribution of datapoints suggests that as the proportion of fat-free mass increases in the sample (moving towards the peak of the triangle), the proportion of truncal and non-truncal fat decrease in almost equal proportions. The main variability is between %fat-free mass and both truncal and non-truncal fat. At a constant percentage of fat-free mass, the variation between truncal and non-truncal fat is at most about 5%. At low proportions of fat-free mass (<50%), the datapoints veer towards non-truncal fat, suggesting that those with low %fat-free mass have relatively more %non-truncal fat compared with %truncal fat. In other words, people with a lot of fat tend to store fat in their limbs rather than on their trunk, possibly because their trunk fat stores are already “full”.

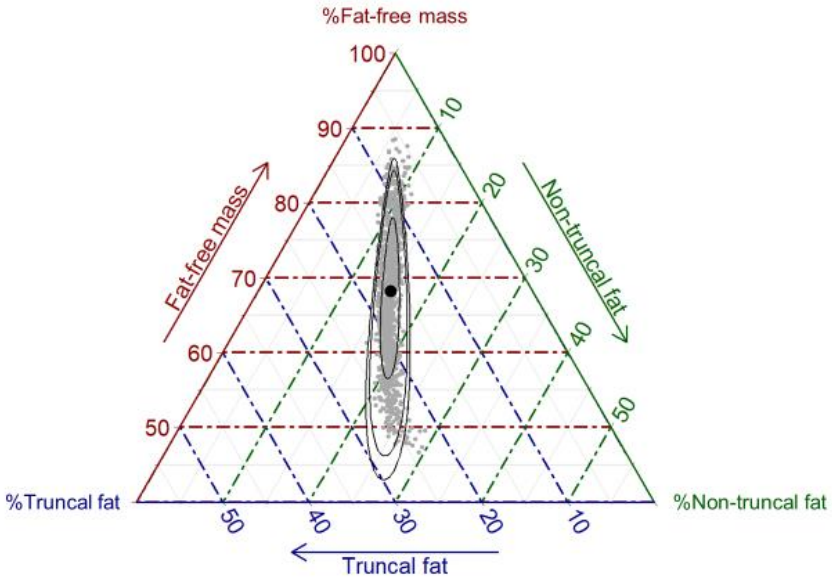

Figure 1. Relative information for body composition among n=1090 adults. Black dot = compositional centre, surrounded by 50, 90 and 95% Gaussian prediction ellipses.

The almost equal variation in %truncal and %non-truncal fat is reflected in the variation matrix (Table 2), which is calculated as the variance in the logratio of two parts. The closer to zero the variance is, the more proportional the values in the two parts are. The variation matrix value for %truncal fat and %non-truncal fat is only 0.01, much smaller than the other values of 0.19 and 0.18.

Table 2. Body composition variation matrix

	Truncal fat	Non-truncal fat	Fat-free mass
Truncal fat	0		
Non-truncal fat	0.01	0	
Fat-free mass	0.19	0.18	0

Values are calculated as the variation in the logarithm of one body part relative to one other body part. Large variances indicate high differentiation in variation between the two parts (low-proportionality), whereas low variances indicate high proportionality between parts. Zero variance indicates perfect proportionality, i.e., one part varies exactly the same as the other.

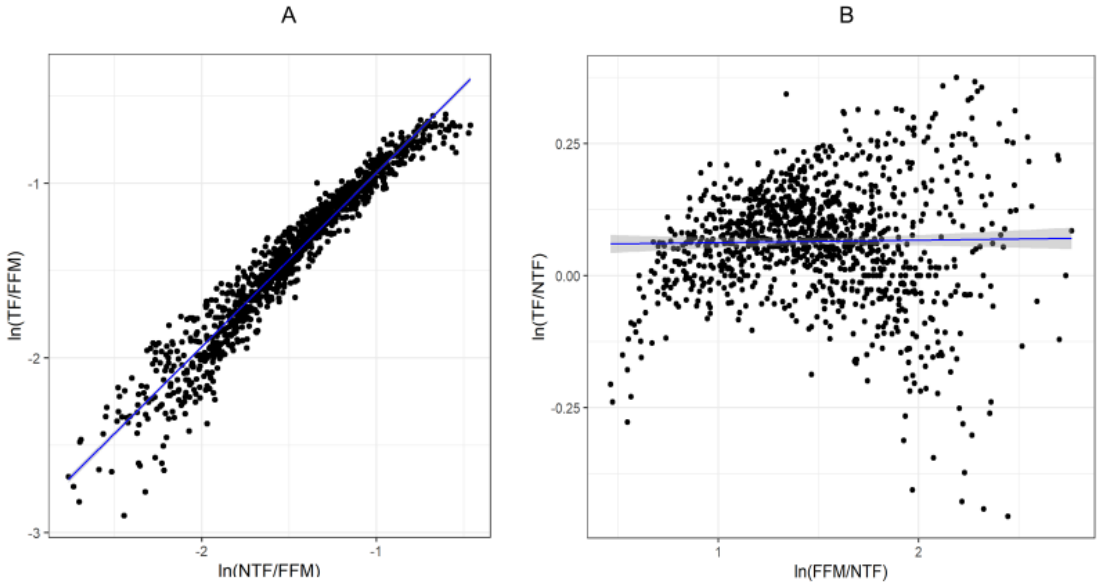

Figure 2. Body composition ALRs. Linear regression lines shown in blue. FFM=fat-free mass; NTF=non-truncal fat; TF=truncal fat.

The high proportionality between %truncal and %non-truncal fat can be visualised by plotting pair-wise logratios against each other (Figure 2).

Figure 2A shows that as truncal fat increases relative to fat-free mass, non-truncal fat also increases (almost proportionately) in relation to fat-free mass. The flattening at the top suggests an upper limit to %truncal fat (about 50% of body mass). This may be because the trunk must contain a lot of non-fat material, e.g., the axial skeleton, the organs, most of the blood and stomach contents. The proportionality between %truncal and %non-truncal fat can also be seen by the relatively flat regression line in Figure 2B. As the proportion of fat-free mass increases relative to non-truncal at, there is no clear difference in the proportion of truncal vs non-truncal fat.

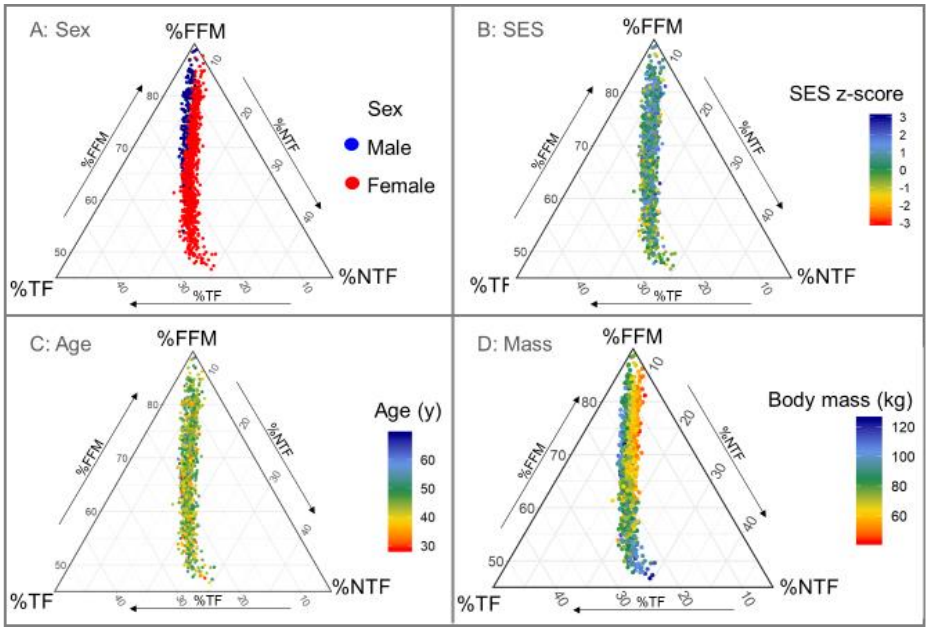

Figure 3. Exploration of patterns in body composition data. Panel A:Sex distribution; Panel B: SES distribution, N.B., scale truncated at -3 at the lower end as only eight observations had a *z*-score of <-3; Panel C: Age distribution; Panel D: Body mass distribution. FFM=fat-free mass; NTF=non-truncal fat; TF=truncal fat; SES=socioeconomic status.

Figure 3A suggests that there are distinct sex differences in body composition; females (red) have higher proportions of non-truncal fat, and lower proportions of fat-free mass compared with males (blue). Note, this needs to be considered with caution as there were only 151 (14%) males in the sample (Table 1). Figure 3D shows that people with low body mass appear to have lower proportions of truncal fat and higher fat-free mass than those with high body mass. This suggests that as adults put on weight, it is linked with an increase in fat. There do not appear to be clear differences according to socioeconomic status (Figure 3B) or age (Figure 3C), however most participants were of high socioeconomic status and of younger age with little variation in the sample (Table 1).

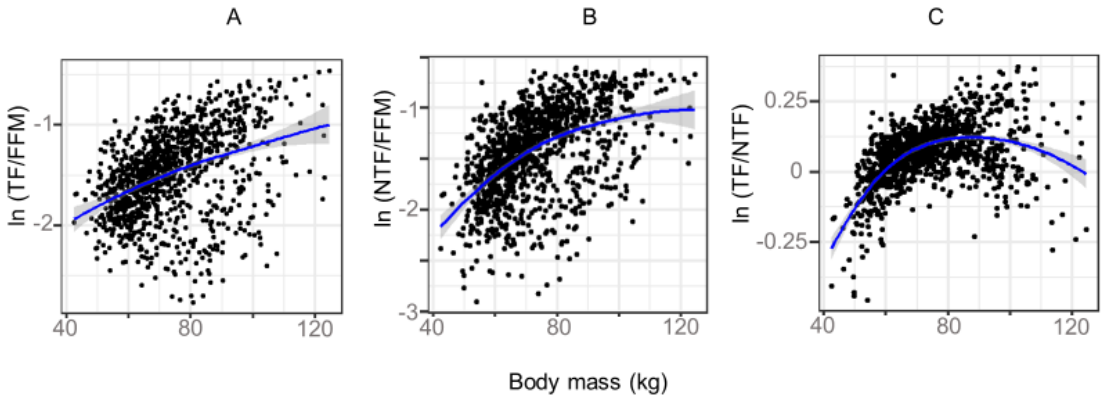

Figure 4. Body composition of pairwise logratios against body mass (kg). TF=truncal fat; FFM=fat-free mass; NTF=non-truncal fat. Blue line represents fitted loess curve.

Plots of pairwise logratios against body mass (Figure 4) show that with increasing body mass, truncal fat and non-truncal fat increase relative to fat-free mass (Plot A and B). At low body mass (<60 kg), on average the non-truncal fat exceeds that of truncal fat (Plot C). The two types of fat are balanced at about 60 kg body mass, after which non-truncal fat exceeds truncal fat. As body mass increases, truncal fat increases relative to non-truncal flat, but the relationship appears to plateau and becomes less clear as there are fewer data for those with body mass >100 kg.

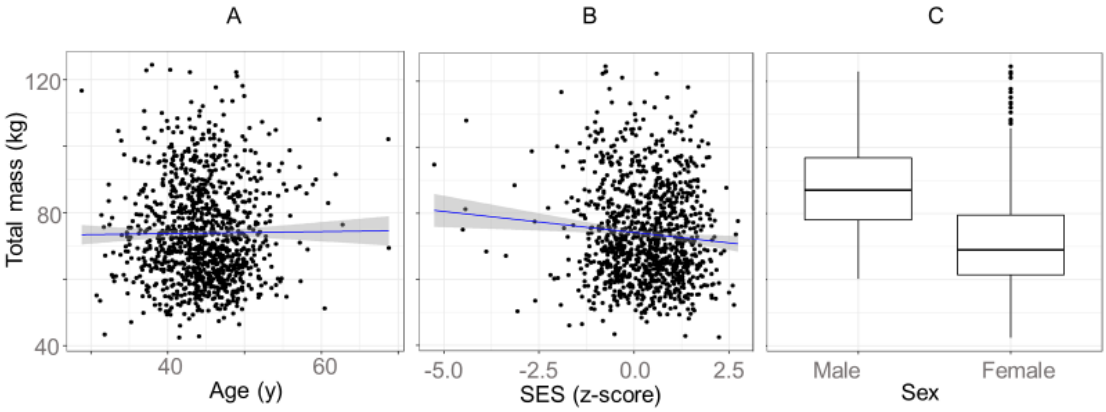

Figure 5. Body mass (kg) against age, socioeconomic status (SES *z*-score) and sex. Blue line in Panels A and B represents fitted loess curve.

Body mass and age do not show any clear relationship (Figure 5A), while there appears to be a slight negative relationship between mass and socioeconomic status (Figure 5B). It can be clearly seen that males tend to have higher body mass than females (Figure 5C).

4.3 Relationship between body composition, body mass and bone strength

Exploratory plots suggest body composition is associated with bone strength. In a ternary diagram (Figure 6), better bone strength (blue) is observed at higher proportions of fat-free mass (>65%) where %truncal fat (5-25%) exceeds %non-truncal fat (5-15%). The best bone strength zone (circled) coincides with the distribution of male participants, and the area of highest body mass (Figure 3). In the analysis of body composition against bone strength it would seem important to account for sex and body mass, as better bone strength at high proportions of fat-free mass may be due to being male and/or having a higher body mass.

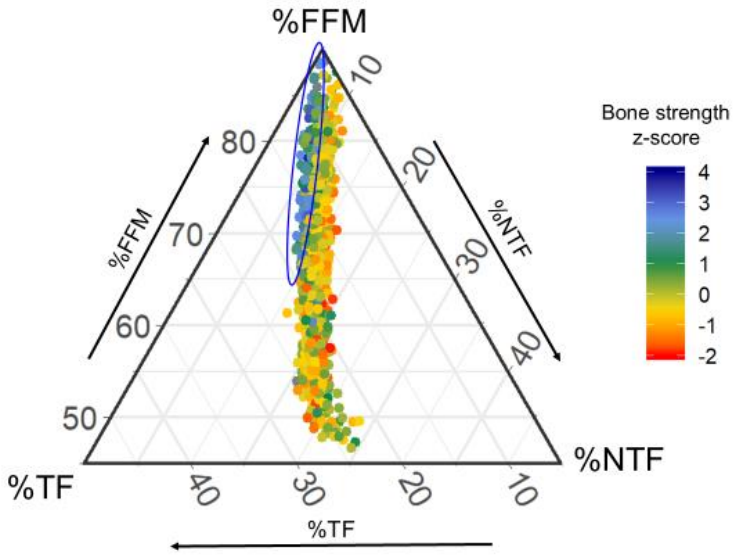

Figure 6. Distribution of bone strength *z*-score across body composition data. FFM=fat-free mass; TF=truncal fat, NTF=non-truncal fat.

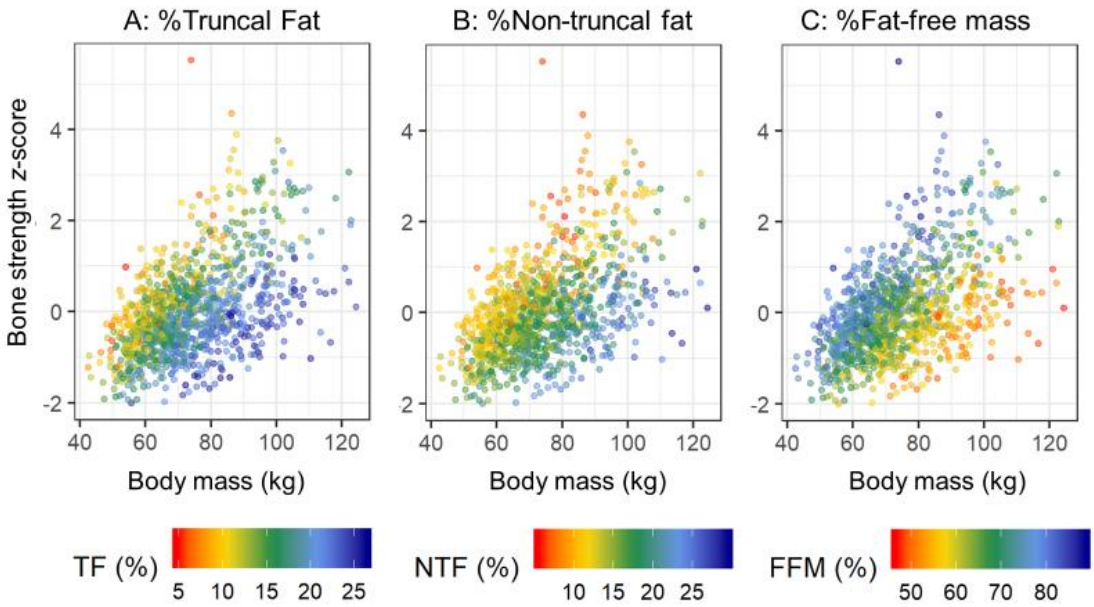

Figure 7. Bone strength vs body mass (kg), as observed in the sample. Datapoints coloured according to body composition; Panel A: %TF (Truncal fat); Panel B: %NTF (Non-truncal fat); Panel C: %FFM (Fat-free mass).

The scatter plot of body mass against bone strength scores show that as body mass increases, bone strength also increases (Figure 7). In each panel, the datapoints are coloured according to the percentage of a different part of the body composition. The figure suggests that whatever the body mass, lower proportions of fat (red colours, Figure7A and B) are associated with better bone strength, while higher proportions of fat-free mass (blue colour, Figure 7C) are associated with better bone strength. One datapoint appears to be an outlier, having very high, but still feasible, bone strength (*z*-score > 5). The impact of this outlying observation will require consideration in subsequent analyses.

Regression models

The relationships between body composition, body mass and bone strength were explored using the tCoDA approach. In this section we demonstrate both the interpretation of individual regression coefficients (using ILRs) and model-based prediction (using ILRs or ALRs). It is important to stress that due to the cross-sectional nature of this study's data, the ensuing analysis explore associations and do not attempt to infer causality. We describe potentially causal relationships which are plausible in real life, but cannot be easily tested in controlled experimental trials

because they would require invasive treatments such as bariatric surgery or liposuction.

Using linear regression models, body composition logratios and the multiplicative total term were regressed against bone strength z-score, with adjustment for sex, age and socioeconomic status. Initial models included interaction terms between sex and body composition/body mass, but these were dropped from the final models as they were insignificant. Interaction between body composition and body mass was significant but in this case it was dropped as it did not improve the predictive performance of the model. Model performance was evaluated using 10-fold cross-validation (caret[23] package in R). The average R-squared and prediction error rate of the ten iterations are presented in Table 3.

Table 3. Linear regression models: prediction error rates from 10-fold cross validation; mean (SD).

Independent variables	R squared	Root mean squared error	Mean absolute error
ILRs only	23.51 (7.57)	87.71 (7.38)	68.83 (4.29)
ILRs and total	64.10 (5.46)	59.93 (4.7)	46.56 (3.11)
ILRs, total and covariates	65.29 (4.28)	59.10 (5.35)	46.40 (3.12)
ILRs, total and covariates, and interaction between ILRs and total	65.28 (4.73)	59.03 (5.28)	46.10 (3.25)

Values are expressed in percentages. ILR = isometric logratios.

4.4 Interpretation of ILR coefficients

The ILRs were created using a SBP that divided parts into groups of interest. The SBP divided fat from non-fat in the first ILR and divided between fat parts in the second ILR (Table 4).

Table 4. Specifically selected Sequential Binary Partition (SBP)

	Truncal fat	Non-truncal Fat	Fat-free mass
ILR1	–1	–1	+1
ILR2	+1	–1	0

ILR = isometric logratio.

The following ILRs were created from the SBP in Table 4: $ILRs = \left[\sqrt{\frac{2}{3}}\ln\frac{FF}{\sqrt{TF \cdot NTF}}, \sqrt{\frac{1}{2}}\ln\frac{TF}{NTF}\right]$, where FF = fat-free mass, TF = truncal fat and NTF = non-truncal fat.

The set of ILRs was used to represent body composition in the following multiple linear regression model which also included the recommended multiplicative term for absolute body mass:

$$\dot{zBoneStrength} = \dot{\beta}_0 + \dot{\beta}_1\sqrt{\frac{2}{3}}\ln\frac{FF}{\sqrt{TF \cdot NTF}} + \dot{\beta}_2\sqrt{\frac{1}{2}}\ln\frac{TF}{NTF} + \dot{\beta}_3\frac{1}{\sqrt{3}}\ln(TF \cdot NTF \cdot FF) + sex + age + SES.$$

The set of ILRs (i.e., body composition) was significantly associated with bone strength (MANOVA statistic for the set of logratios: *Sum Sq* = 87.7, *F* = 126.8, $p < 0.001$). Model parameters are shown in Table 5.

Table 5. Model parameters of the ILR tCoDA regression model for the relationship between body composition and bone strength

Term	Beta	SE	Standardised beta	*t*	*p*
(Intercept)	-16.04	0.66	0.00	-24.4	<0.001
ILR1 (FFM vs remaining)[a]	3.48	0.13	1.22	26.4	<0.001
ILR2 (TF vs NTF)	-0.75	0.29	-0.06	-2.6	0.009
Multiplicative total[b]	2.35	0.09	1.20	27.4	<0.001
Age (y)	0.01	0.00	0.04	2.3	0.02
Sex (Female)	-0.38	0.08	-0.13	-4.6	<0.001
z-SEP	0.04	0.02	0.04	2.3	0.02

ILR = isometric logratio; tCoDA = Compositional Data Analysis with a total; FFF=fat-free mass; TF=truncal fat; NTF=non-truncal fat; SEP = socioeconomic position. [a]ILR defined as $ILR = \left[\sqrt{\frac{2}{3}}\ln\frac{FFM}{\sqrt{TF \cdot NTF}}, \sqrt{\frac{1}{2}}\ln\frac{TF}{NTF}\right]$. [b]Total defined as $t = \frac{1}{\sqrt{3}}\ln(TF(kg) \cdot NTF(kg) \cdot FFM(kg))$. Model adjusted R-squared = 0.65.

The beta estimate for ILR1 (Table 5) indicates that as fat-free mass increases at the expense of truncal fat and non-truncal fat (in equal proportions), bone strength increases ($standardised\ \dot{\beta}$=1.22, $p < 0.001$). As described in Coenders et al,[14] this

beta is associated with the effect of multiplying fat-free mass by a constant (a > 1), while simultaneously multiplying truncal and non-truncal fat by the constant $\frac{1}{\sqrt{a}}$. This is how ILR2 and the total term can be kept constant while ILR1 is varied. The differences in body composition represented by ILR1 (exchanging fat-free mass for both truncal and non-truncal fat, equally) appears to be the most commonly observed situation in this sample.

The beta estimate for ILR2 ($standardised\ \dot{\beta}$=-0.06, p < 0.009) indicates that bone strength decreases as truncal fat increases at the expense of non-truncal fat. Specifically, this beta is associated with the effect of multiplying truncal fat by a constant (a > 1), while multiplying non-truncal fat by $\frac{1}{a}$. In this way, ILR1 and the total term are kept constant while ILR2 is varied. This situation may not commonly occur in the population sampled in this study, as %truncal and %non-truncal fat tend to co-vary in the same direction (Figure 1, Figure 2 and Table 2).

Another option is to explore the relationship between bone strength and each of the parts, relative to the remaining parts. The ILR1 (FFM vs remaining) created by the SBP from Table 4 represents fat-free mass, relative to the remaining parts. Other SBPs are needed to create ILR1s that represent truncal fat (Table 6) and non-truncal fat (Table 7), each relative to the remaining parts.

Table 6. Sequential Binary Partition (SBP) for truncal fat, relative to remaining parts

	Truncal fat	Non-truncal Fat	Fat-free mass
ILR1 (TF vs remaining)	+1	–1	–1
ILR2	0	+1	–1

ILR = isometric logratio; TF = truncal fat.

Table 7. Sequential Binary Partition (SBP) for non-truncal fat, relative to remaining parts

	Truncal fat	Non-truncal Fat	Fat-free mass
ILR1 (NTF vs remaining)	–1	+1	–1
ILR2	–1	0	–1

ILR = isometric logratio; NTF = non-truncal fat.

The two sets of ILRS created by the SBPs from Table 6 and 7 were regressed against bone strength z-score in two further linear models. The multiplicative total and covariates were also included in the models. Table 8 shows the beta estimates for ILR1 from all three regression models, where ILR1 represented one part relative to the remaining parts. Note the first row of Table 8 for ILR1 (FFM vs remaining) is identical to the second row in Table 5. Regression coefficients for the multiplicative total and covariates are identical across all the models, and can be found in Table 5. The findings suggest %fat-free mass is most beneficially associated with better bone strength, and that %truncal fat appears more detrimentally associated than %non-truncal fat (when absolute mass is kept constant).

Table 8. Parameters from tCoDA regression models on bone strength

Model				Standardised		
	Isometric logratio	Beta	SE	beta	*t*	*p*
Model 1	ILR1 (FFM vs remaining)	3.48	0.13	1.22	26.4	<0.001
Model 2	ILR1 (TF vs remaining)	-2.39	0.26	-0.46	-9.1	<0.001
Model 3	ILR1 (NTF vs remaining)	-1.09	0.25	-0.20	-4.3	<0.001
From all Models	Multiplicative total	2.35	0.09	1.20	27.36	<0.001

ILR1 = first isometric logratio; TF = truncal fat, NTF = non-truncal fat; FF = fat-free mass. Models adjusted for age, sex and socioeconomic position.

Interpretation of coefficient for multiplicative total

The beta coefficient for the multiplicative total can be interpreted as the estimated association of changing absolute total multiplicative mass by one unit, whilst keeping body composition constant. While this scenario may be unexpected in empirical situations among adults, the hypothetical situation assists us to disentangle the influence of absolute and relative information. It may also be of interest for cross-sectional comparisons and in younger populations where body mass may be more likely to increase due to skeletal growth rather than fat deposition.

The beta for multiplicative total body mass was positive and significant (*standardised* $\dot{\beta} = 1.20$, $p < 0.001$; Table 5), suggesting that if body mass were to increase independent of changes in body composition, this would be associated with greater

bone strength. The value of the beta for body mass cannot be interpreted as estimated change in bone strength when body mass is increased by one kg, as the term for body mass is derived by multiplying together the values of absolute mass in each part. Instead, the beta estimate for body mass represents the association of multiplying absolute values (kg) of all body parts (truncal fat, non-truncal fat and fat-free mass) simultaneously with the same constant ($a > 1$). As each part is multiplied by the same constant, relative information (body composition) remains unchanged, meaning the logratios can be kept constant as body mass varies.

Diagnostic checks of ILR regression model

Diagnostic checks of the model fit suggested the linear multiple regression model provided a good fit to the data (Figure 8). As noted earlier, observation 15 appeared to be an outlier. The observed bone strength score for this participant was very high, but the Residuals vs Leverage plot suggested the influence of this observation was within acceptable limits, so the participant was retained for the analyses.

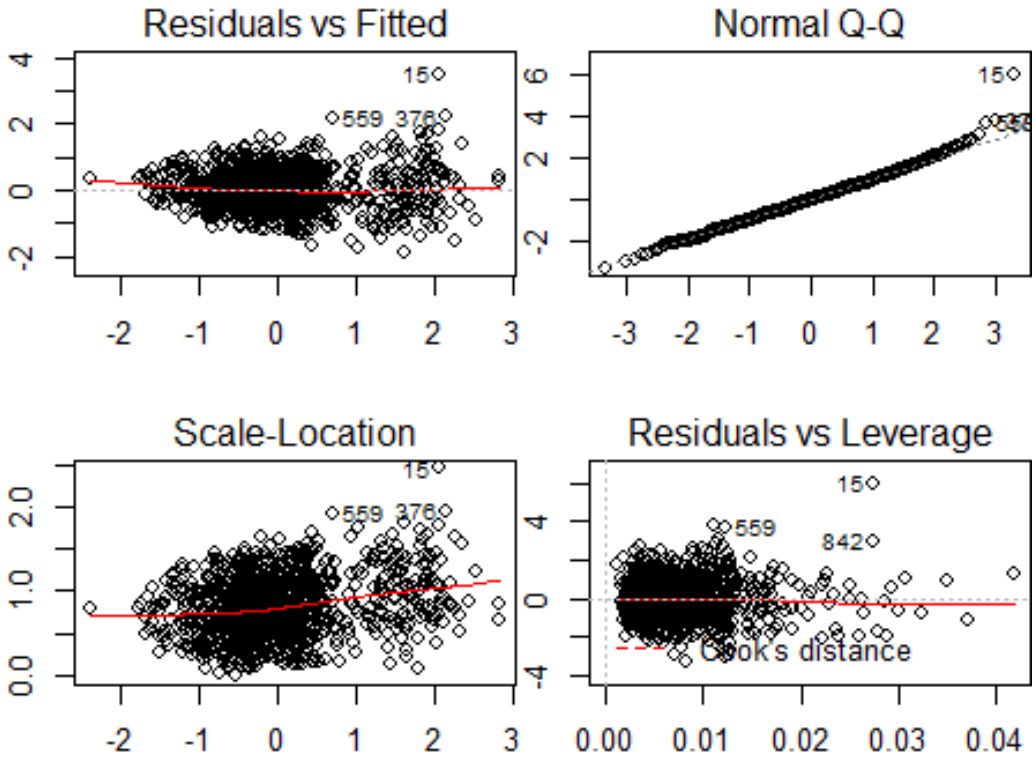

Figure 8. Model diagnostic plots for linear regression model.

Exploration of the variance inflation factors of the four ILR models revealed very high values for the first ILRs and the multiplicative total (generalized variance inflation factors (VIFs) for ILR1 (TF vs remaining) = 8.1; ILR1 (NTF vs remaining) = 6.5; ILR1 (FFM vs remaining) = 6.7; multiplicative total = 6.0).

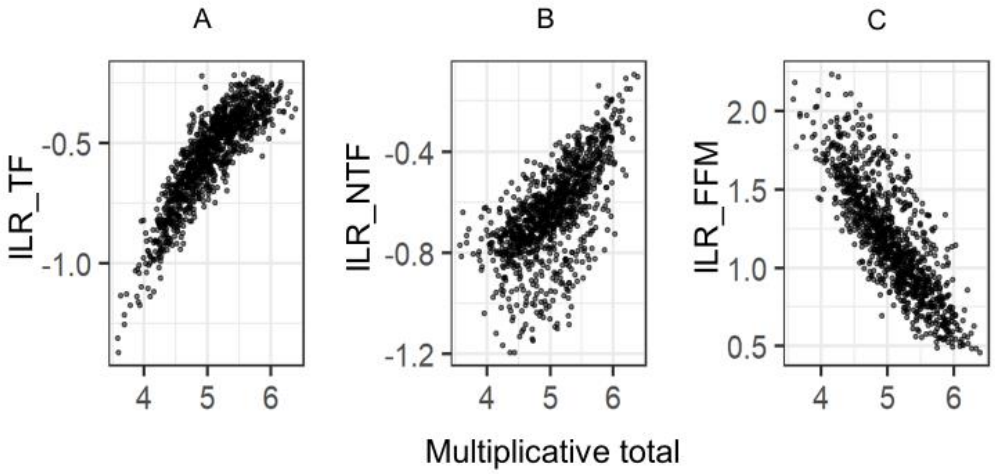

Figure 9. Correlation between the first ILR and multiplicative total. ILR=isometric logratio. ILR_TF=truncal fat vs remaining parts; ILR_NTF=non-truncal fat vs remaining parts; ILR_FFM=fat-free mass vs remaining parts.

Correlation between the multiplicative total and the first ILR from each set of coordinates was high, particularly for the ILR of truncal fat vs remaining (Pearson's r = 0.87, $p<0.001$) and fat-free mass vs remaining (Pearson's r = -0.82, $p<0.001$) (Figure 9), consistent with Fig. 3D.

Co-dependency between explanatory variables may lead to inflated variability in the beta coefficients (as reflected in the model VIFs), thus the interpretation of the regression coefficients for the first ILR and multiplicative total should be considered with caution. However, it is reassuring that the regression coefficients corroborate with patterns observed in our exploratory analyses (Figures 3D and 7). As models with high collinearity between explanatory variables can still perform well for prediction purposes, for this particular sample a predictive approach may be better

suited to explore the relationships between body composition, body mass and bone strength.

4.5 Predictive approach

In this section, we use the tCoDA model (with ALRs or ILRs) to predict the associated difference in bone strength for a number of hypothetical situations. Firstly, we predict bone strength for varying body composition, without changing the absolute body mass. Secondly, we predict bone strength for varying body mass, without changing body composition. These two situations, although interesting for certain populations, may not reflect common circumstances for this population. In our final hypothetical situation, we predict bone strength for varying body mass *and* varying body composition, based on their association in the sample.

Varying body composition, constant body mass

We estimated the difference in bone strength associated with the substitution of 5 percentage units of body mass between two body compartments (one-for-one isocompartmental substitution). The results from isocompartmental substitution allow the estimated effects of exchanging the same absolute proportion of mass from one part to another to be evaluated side-by-side. The substitution analyses was done by predicting bone strength for two scenarios: (1) average body composition, and (2) a body composition where 5% had been taken from one part (e.g., truncal fat) and given to another (e.g., fat-free mass), and (3) finding the difference between the two bone strength predictions. Body mass and all covariates were kept constant in the prediction models. Findings for all possible 5% one-for-one isocompartmental substitutions are in Table 9.

Note, the situation where 5% is reallocated between some body composition parts (e.g., %truncal and %non-truncal fat while keeping %fat-free mass constant at the mean value) is at the extremes of what might be feasible or physiologically possible in this population of adults (see Figure 1). A more commonly observed reallocation may be between %fat-free mass and both %truncal and non-truncal fat (equally), and predictions could also be calculated for this situation using the same method. However, one-for-one reallocations of mass may be expected in other populations

(growing children, elderly with increasing frailty). In any case, our cross-sectional data do not provide information on within-person reallocations of mass. Longitudinal data are required to model how changes in body composition and body mass are associated with health-related outcomes. Instead, our analyses model population-level differences.

Table 9. Difference in estimated bone strength z-score when 5% units of mass are reallocated between body parts

		95% Confidence Interval	
Substitution	Difference	Upper	Lower
↑TF, ↓NTF	-0.1	-0.3	0.0
↑TF, ↓FF	-0.7	-0.8	-0.6
↑NTF, ↓TF	0.4	0.3	0.5
↑NTF, ↓FF	-0.5	-0.6	-0.4
↑FF, ↓TF	0.8	0.7	0.9
↑FF, ↓NTF	0.5	0.4	0.6

TF = truncal fat, NTF = non-truncal fat, FF = fat-free mass. Substitutions of 5% body mass are relative to the mean body composition (%TF = 16.4, %NTF = 15.4, %FF = 68.3). Body mass kept constant. Analyses additionally adjusted for age, sex, and socioeconomic status.

The largest differences (up to 0.8 SD) in bone strength were associated with substitutions between fat-free mass and truncal fat. This is congruent with inference from the ILR regression coefficients, suggesting fat-free mass and truncal fat are most important in the association between body composition and bone strength (Table 8). The results also suggest that adding %non-truncal fat at the expense of %truncal fat would be associated with stronger bones, even though in practice the two usually move more or less synchronously in the same direction.

The model was used to predict bone strength over the body composition values observed among the participants. The total term and all covariates were kept constant. The predicted bone strength scores were plotted on a ternary surface, superimposed over their predictive body composition. The values of predicted bone strength were colour-coded along a heat-gradient so the weakest bones were red and the strongest

bones were blue. Figure 10 presents the association between body composition and bone strength, independent of body mass.

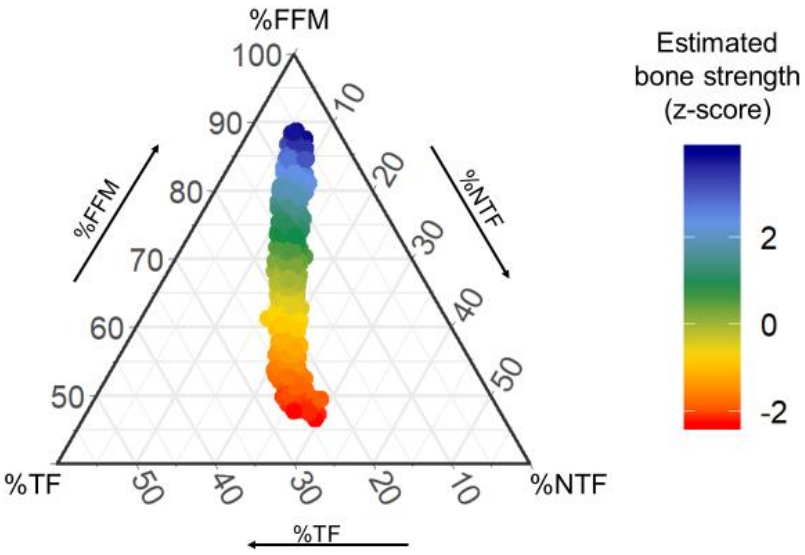

Figure 10. Response surface of predicted bone strength *z*-score overlaid on body composition ternary plot (body mass kept constant). Analyses additionally adjusted for age, sex and socioeconomic status. Best (blue) and worst (red) body compositions for bone strength are identified.

The gradient of response surface in Figure 10 suggests that at any given body mass, the relationship between bone strength and body composition is predominantly driven by the proportion of fat-free mass relative to both truncal and non-truncal fat. The higher the proportion of fat free mass (and lower non-truncal and truncal fat), the better the bone strength. The gradient is slightly skewed toward the right bottom vertex of the triangle, suggesting the proportion of truncal fat is more strongly associated than the proportion of non-truncal fat. Within the range of sampled body compositions, there were about 5 standard deviations difference in estimated bone strength between the best and worst body compositions.

Varying body mass, constant body composition

The same regression model was used to predict bone strength for varying body mass (in 10 kg increments) expressed as the multiplicative total, keeping body composition and covariates constant. A response curve was generated by smoothing between bone strength predicted for incremental differences in overall body mass (Figure 11). Note,

there were few observations at body masses >100kg, meaning we have less certainty in these predictions.

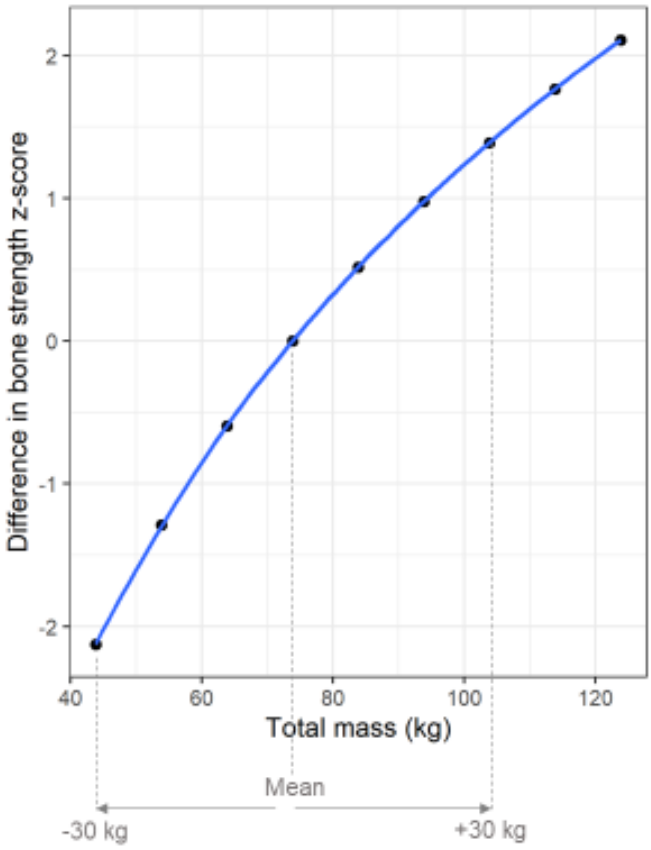

Figure 11. Estimated difference in bone strength when absolute body mass is changed around the mean body mass (keeping body composition constant). Analysis additionally adjusted for age, sex, and socioeconomic position.

Varying body composition and body mass

In real life, changes in body composition may not be independent of changes in body mass, so the scenarios modelled above may not represent commonly occurring situations. To explore how body mass (kg) was associated with body composition in this sample, body composition logratios (outcomes) were regressed against the summative total (kg) (explanatory). Diagnostic plots showed that a quadratic term for body mass was a better fit to the data, which was confirmed by lower RMSEs for the quadratic models compared to the linear models. The null hypothesis of no relationship was rejected, as body mass was a significant predictor of the set of logratios (MANOVA *SumSq*=0.51, *F*=185, $p<0.001$). This suggests the situation

modelled above (increasing body mass while keeping the body composition constant) is unlikely to happen in real life for people represented by this sample (i.e., predominantly mothers, mean age of 45 years (SD=5, range=29 – 69)).

To describe how body composition was associated with body mass we used the multivariate regression model to predict body composition logratios for a sequence of body masses ranging from 40-120 kg. The predicted logratios were expressed in percentage units (Table 10) and plotted relative to the sample mean body composition (Figure 12).

Table 10. Predicted body composition for increasing values of body mass

Body mass (kg)	%Truncal Fat	% Non-truncal fat	% Fat-free mass
40	8	11	81
50	11	13	76
60	14	14	72
70	16	15	69
80	18	16	66
90	19	17	63
100	20	18	61
110	21	19	60
120	21	21	58

Body composition estimated from regression model with squared term for body mass (log-transformed).

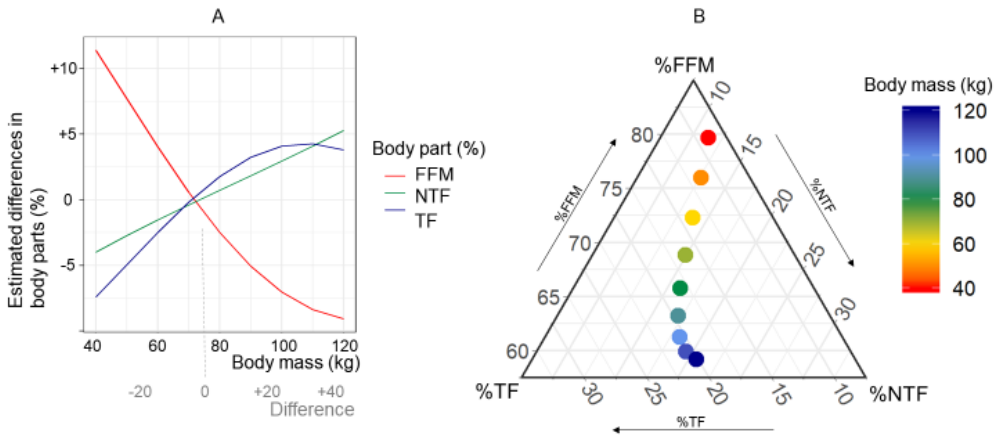

Figure 12. Estimated body composition for varying body mass (squared term). TF=truncal fat; NTF=non-truncal fat; FFM=fat-free mass.

As body mass increases in the sample, %fat-free mass tends to be replaced with %truncal fat and to a slightly lesser extent, %non-truncal fat. At high body masses (>100 kg), there is flattening of the estimated differences in body composition (Figure 12A) and datapoints are closer together (Figure 12B), indicating that the hypothetical situation of changing body mass whilst keeping body composition constant may be a reasonable representation among those with high body mass. However, as previously mentioned, caution must be exercised when interpreting results at high masses as there were few observations >100 kg. Using the predicted body compositions (expressed as logratios) from Table 10, and the corresponding body mass (expressed as the multiplicative total), we used the original regression models to estimate bone strength (adjusted for age, sex and socioeconomic status). Thus, both relative and absolute information were varied.

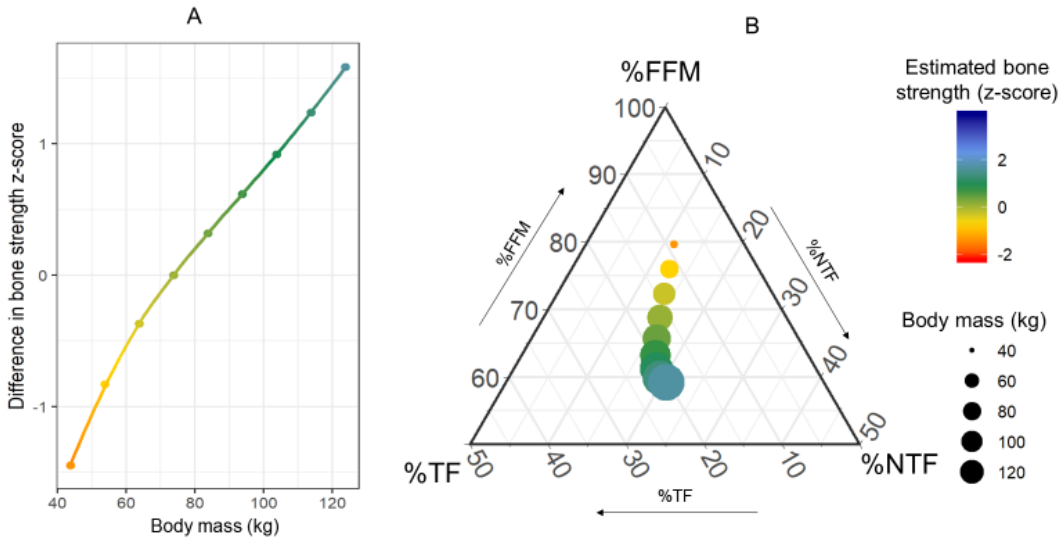

Figure 13. Plot A: Estimated bone strength for varying body masses, at population-estimated body compositions. Plot B: Estimated bone strength for varying body compositions, at population-estimated body mass. All analyses adjusted for age, sex and socioeconomic status. FFM=fat-free mass; TF=truncal fat; NTF=non-truncal fat.

At a glance, the relationship between body composition and bone strength (plotted in Figure 13B) appears to contradict the relationship shown in Figure 10. This is because estimates in Figure 10 assume a constant body mass at the sample mean value, but body mass varies in Figure 13. When body mass is allowed to vary along with body composition, the positive relationship between body mass and bone strength is reflected in Figure 13B, and masks the relationship between body composition and bone strength. Body composition is still associated with bone strength (%fat-free mass positively, and %non-truncal and %truncal fat negatively) because the positive association with body mass is largely attenuated in Figure 13A when compared with Figure 10 (where body composition is kept constant at the mean value).

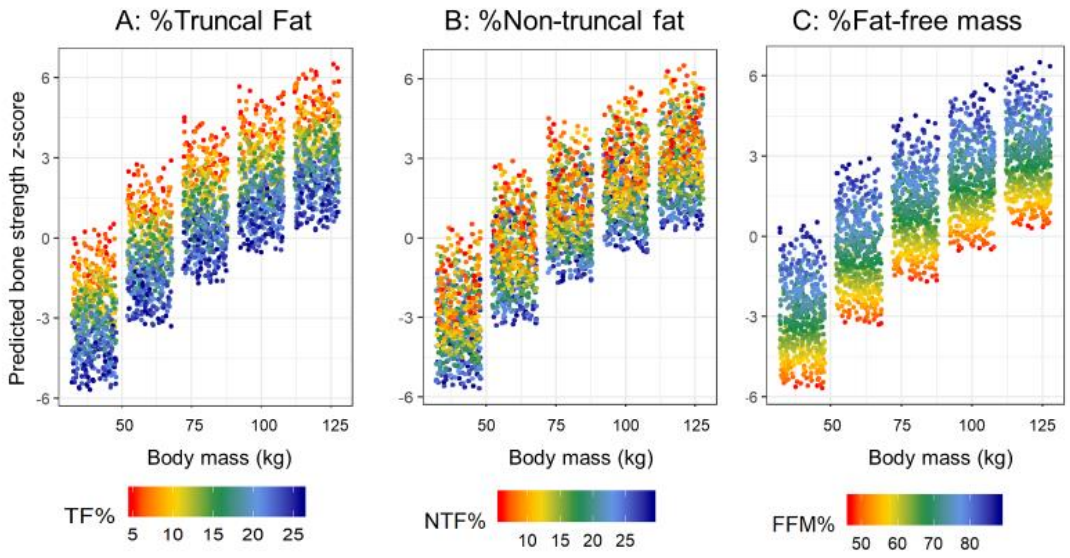

Figure 14. Relationship between body mass and bone strength, and the association of body composition across body mass. Panel A is coloured according to %truncal fat (TF); Panel B according to %non-truncal fat (NTF) and Panel C according to %fat-free mass (FFM).

The relationship between body composition at each level of body mass can be visualised by plotting estimated bone strength for every possible body composition at 40, 60, 80, 100 and 120 kg (Figure 14). A very clear positive gradient can be seen for %fat-free mass and bone strength, and a negative gradient for %truncal fat. There is also a negative gradient for %non-truncal fat but it is less clear, suggesting that the

relationship with bone health is more closely related to the values of %fat-free mass and %truncal fat than %non-truncal fat.

5. Comments

Body composition is a widely studied exposure in health research. To date, studies have largely analysed the health associations of each part separately. We present an approach (tCoDA) that enables all body parts to be included in the same model simultaneously, and that takes into account body mass. The tCoDA approach allows exploration of associations between health and body composition independent of body mass, and of health with body mass independent of body composition. This makes it possible to disentangle relative and absolute information, and to better understand how body composition and overall mass are associated with health outcomes.

The analyses suggest that if the findings were to reflect causal relationships, body mass may be a key determinant of adults' bone strength. This finding is supported by the mechanostat model.[24] The model predicts that increasing overall body mass will induce an adaptive response in the bones, which are continuously remodelling. Bones will respond to increased mechanical loads by increasing bone mass and improving the quality of bone microarchitecture.

We found that although higher mass was associated with better bone strength, the body composition also had important associations with bone strength. At any given body mass, higher proportions of fat-free mass relative to truncal and non-truncal fat were associated with better bone strength. This may be because higher fat-free mass likely represents relatively more muscle mass, meaning bones are being loaded by muscle action (functional strain) as well as static strain from body mass, leading to stronger bones (mechanostat hypothesis). The skeletal load from fat is limited to static strain. In addition, higher proportions of fat may influence endocrine function, leading to dysregulation of bone formation.[24] In relation to the distribution of body fat, we found the associations of %truncal fat to be slightly more detrimental to bone strength than those of %non-truncal fat, but both these parts appeared to vary together, i.e., those with high %truncal fat also had high %non-truncal fat.

While our interest is in exploring associations for which causality is plausible and would be important, our cross-sectional data cannot confirm cause and effect. We adjusted our models for several potential confounders, however there still may be residual confounding due to other factors, for example, the 24-hour activity composition (physical activity, sedentary time and sleep). Future work could explore the application of graphical models (directed acyclic graphs) and counterfactual approaches to estimate causal effects.

We describe in this paper the basic principles of compositional data analysis, specifically tCoDA, for body composition data linked with health outcomes. There is an opportunity to further explore other applications of compositional data analysis for body composition data, such as cluster analysis, non-linear models and considering body composition as an outcome. The example presented in this paper used a 3-part body composition operationalized by distribution of fat, but other ways of compartmentalising body mass may be of interest (e.g., skeletal, muscle and visceral tissue). Future work may explore means of visualising body compositions with more than three parts, and how to account for potential sources of error (such as measurement uncertainty) in the analysis. Our main goal is to motivate research in the application of compositional data analysis in body composition research.

The tCoDA approach enables a comprehensive exploration of the associations between body composition, body mass and health outcomes. Its application has the potential to provide sound evidence to advise health interventions and guide public health policy.

ACKNOW

J.A. Martín-Fernández was supported by the Spanish Ministry of Science, Innovation and Universities under the project CODAMET (RTI2018-095518-B-C21, 2019-2021).

References

1. Bello NA, Cheng S, Claggett B, et al. Association of weight and body composition on cardiac structure and function in the ARIC study (Atherosclerosis Risk in Communities). *Circulation: Heart Failure.* 2016;9(8):e002978.
2. Koster A, Stenholm S, Alley DE, et al. Body fat distribution and inflammation among obese older adults with and without metabolic syndrome. *Obesity.* 2010;18(12):2354-2361.
3. Shapses SA, Cifuentes M. Body weight/composition and weight change: effects on bone health. In: *Nutrition and Bone Health.* Springer; 2015:561-583.
4. Santanasto AJ, Goodpaster BH, Kritchevsky SB, et al. Body composition remodeling and mortality: the health aging and body composition study. *The Journals of Gerontology: Series A.* 2017;72(4):513-519.
5. Chang S-H, Beason TS, Hunleth JM, Colditz GA. A systematic review of body fat distribution and mortality in older people. *Maturitas.* 2012;72(3):175-191.
6. Kim J, Kwon H, Heo B-K, et al. The Association between fat mass, lean mass and bone mineral density in premenopausal women in Korea: A cross-sectional study. *Korean Journal of Family Medicine.* 2018;39(2):74.
7. Rokoff LB, Rifas-Shiman SL, Switkowski KM, et al. Body composition and bone mineral density in childhood. *Bone.* 2019;121:9-15.
8. Kang SM, Yoon JW, Ahn HY, et al. Android fat depot is more closely associated with metabolic syndrome than abdominal visceral fat in elderly people. *PloS one.* 2011;6(11):e27694.
9. Aitchison J. The statistical analysis of compositional data. *Journal of the Royal Statistical Society: Series B (Methodological).* 1982;44(2):139-160.
10. Dumuid D, Wake M, Clifford S, et al. The association of the body composition of children with 24-hour activity composition. *The Journal of Pediatrics.* 2019;208:43-49. e49.
11. Greenacre M. Towards a pragmatic approach to compositional data analysis. *Economics Working Papers 1554* 2017; https://ideas.repec.org/p/upf/upfgen/1554.html.
12. Hron K, Filzmoser P, Thompson K. Linear regression with compositional explanatory variables. *Journal of Applied Statistics.* 2012;39(5):1115-1128.

13. Egozcue JJ, Pawlowsky-Glahn V. Groups of parts and their balances in compositional data analysis. *Mathematical Geology.* 2005;37(7):795-828.
14. Coenders G, Martín-Fernández JA, Ferrer-Rosell B. When relative and absolute information matter: compositional predictor with a total in generalized linear models. *Statistical Modelling.* 2017;17(6):494-512.
15. Dumuid D, Pedišić Ž, Stanford TE, et al. The compositional isotemporal substitution model: A method for estimating changes in a health outcome for reallocation of time between sleep, physical activity and sedentary behaviour. *Statistical methods in medical research.* 2019;28(3):846-857.
16. Mateu-Figueras G, Pawlowsky-Glahn V, Egozcue JJ. The principle of working on coordinates. *Compositional Data Analysis.* 2011:29-42.
17. Pawlowsky-Glahn V, Egozcue JJ, Lovell D. Tools for compositional data with a total. *Statistical Modelling.* 2015;15(2):175-190.
18. Clifford SA, Davies S, Wake M. Child Health CheckPoint: Cohort summary and methodology of a physical health and biospecimen module for the Longitudinal Study of Australian Children. *BMJ Open.* 2019;9(Suppl 3):3-22.
19. Wake M, Clifford S, York E, et al. Introducing growing up in Australia's child health checkpoint: a physical and biomarkers module for the longitudinal study of Australian children. *Family Matters.* 2014;95:15-23.
20. Vlok J, Simm P, Lycett K, et al. pQCT bone geometry and strength: population epidemiology and concordance in Australian children aged 11–12 years and their parents. *BMJ Open.* 2019;9 (Suppl 3):63:74.
21. Clifford S, Gillespie A, Olds T, Grobler AC, Wake M. Popluation epidemiology and concordance in 11-12 year old Australians and their parents. *BMJ Open.* 2019;9 (Suppl 3):95-105.
22. Blakemore T, Strazdins L, Gibbings J. Measuring family socioeconomic position. *Australian Social Policy.* 2009;8:121-168.
23. Kuhn M. caret: Classification and Regression Training. R package version 6.0-85. 2020; https://CRAN.R-project.org/package=caret.
24. Iwaniec U, Turner R. Influence of body weight on bone mass, architecture and turnover. *Journal of Endocrinology.* 2016;230(3):R115–R130.