The fallacy of Ratio Correction to address confounding factors

Natasha A Karp[1], **Anne Segonds-Pichon**[2], **Anna-Karin B Gerdin**[1], **Ramiro Ramírez-Solis**[1], and **Jacqueline K White**[1,*]

[1]Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SA, UK. Tel: +44 (0)1223 834244, Fax: +44 (0)1223 494919

[2]The Babraham Institute, Babraham Research Campus, Cambridge, CB22 3AT, UK. Tel: +44 (0)1223 496000, Fax: +44 (0)1223 496002

Abstract

Scientists aspire to measure cause and effect. Unfortunately confounding variables, ones that are associated with both the probable cause and the outcome, can lead to an association that is true but potentially misleading. For example, altered body weight is often observed in a gene knockout, however many other variables, such as lean mass, will also change as the body weight changes. This leaves the researcher asking whether the change in that variable is expected for that change in weight. Ratio Correction, which is often referred to as Normalization, is a method used commonly to remove the effect of a confounding variable. Although Ratio Correction is used widely in biological research, it is not the method recommended in the statistical literature to address confounding factors; instead regression methods such as the analysis of covariance (ANCOVA) are proposed. This method examines the difference in means after adjusting for the confounding relationship. Using real data, this manuscript demonstrates how the Ratio Correction approach is flawed and can result in erroneous calls of significance leading to inappropriate biological conclusions. This arises as some of the underlying assumptions are not met. The manuscript goes on to demonstrate that researchers should use ANCOVA, and discusses how graphical tools can be used readily to judge the robustness of this method. This study is therefore a clear example of why assumption testing is an important component of a study and thus why it is included in the Animal Research: reporting of *In Vivo* Experiment (ARRIVE) guidelines.

Keywords

Introduction

Scientists aspire to measure cause and effect. Consequently experiments are designed to isolate the effect of interest as we manipulate a variable (the independent variable) to see what effect it has on a second variable (the dependent variable). Unfortunately, confounding variables (also called confounders, lurking variables or confounding factors) can lead to studies where the internal validity can be questioned. A confounding variable is one that it is associated with both the probable cause and the outcome. When confounding variables are uncontrolled, an erroneous conclusion can be drawn. Typically, experiments are designed using random allocation to remove the potential effects of confounding variables. However, in many experiments this is not possible and statistical methods for handling non-equivalent groups are necessary (1). This manuscript uses real data from phenotyping experiments to discuss this issue.

Using the mouse as a model organism, targeted inactivation of a gene can yield a diverse range of phenotypic manifestations (2). Body weight is a highly heritable trait, and consequently commonly altered in knockout lines of mice (3). Body weight is known to correlate with many other biologically interesting variables (e.g. grip strength, heart weight and bone density), therefore it is unsurprising when the knockout also results in difference in these other variables. It raises the question as whether the change in these variables is greater than expected given the observed change in body weight. Statistically, body weight in these examples is described as a confounding variable.

Ratio Correction, which is often referred to as Normalization, is a method frequently used in biological research to remove the effect of a confounding variable (4-9). In this method, data are 'normalised' by the calculation of a ratio with the idea that this will give a relative value. For each individual, the dependent variable measurement is divided by the confounding variable measurement (e.g. grip strength/body weight). A statistical test, such as a Student's *t*-test, is then used to compare the ratio across the control and treatment groups. The theory is that the relative ratio will remove the effect of differences in the confounding variable, as the ratio will be constant if the difference between the variable of interest is due solely to the confounding variable. However, there have been publications, particularly in physiology and anthropology, which have raised concerns about the use of the Ratio Correction approach (10-16). For example, in a discussion of morphometrics data, Reist stated that whilst it was perhaps the most prevalent method, the superficial simplicity of Ratio Correction masked a host of potential problems (14). Similarly, two studies in human obesity looking at variables such as body mass index and percentage body fat, demonstrated that mathematical ratios can lead to spurious conclusions (17, 18).

Statistical literature suggests the use of analysis of covariance (ANCOVA) to address confounding variables. This method examines the difference in means after adjusting for the confounding relationship (19). For example, ANCOVA was used to remove the effect of body weight difference in a study looking at the impact of a weight loss program on coronary heart disease risk factors in mildly obese women (20). Furthermore, the use of ANCOVA was recently recommended by Tschöp and coworkers in a guide to analysis of mouse energy metabolism (21).

This report aims to raise awareness of the generic nature of the issues with Ratio Correction. Two methods, Ratio Correction and ANCOVA, were used to adjust multiple parameters for the effect of confounding variables; the results were compared, with particular focus on mouse phenotyping data. In line with the "Animal Research: reporting of *In Vivo* Experiment" (ARRIVE) guidelines, the assumptions underpinning the analysis methods are assessed (22). The manuscript goes on to recommend that researchers should use ANCOVA, and discusses the strengths and weaknesses of this method.

Materials and Methods

Mice

The care and use of all mice in this study was carried out in accordance with UK Home Office regulations, UK Animals (Scientific Procedures) Act of 1986. Mice were maintained in a specific pathogen free unit on a 12hr light: 12hr dark cycle with lights off at 7:30pm and no twilight period. The ambient temperature was $21 \pm 2°C$ and the humidity was $55 \pm 10\%$. Mice were housed using a stocking density of 3-5 mice per cage (overall dimensions of caging: $(L \times W \times H)$ $365 \times 207 \times 140mm$, floor area $530cm^2$) in individually ventilated caging (Tecniplast Seal Safe1284L) receiving 60 air changes per hour. In addition to Aspen bedding substrate, standard environmental enrichment of two nestlets, a cardboard Fun Tunnel and three wooden chew blocks was provided. Mice were given water and diet *ad libitum*. At 4 weeks of age, mice were transferred from a breeders diet (LabDiets, 5021) to a high fat (21.4% fat by crude content) dietary challenge (Special Diet Services, Western RD 829100).

Dual-energy X-ray absorptiometry (DEXA)

At 14 weeks of age the mice were weighed and then anaesthetized with either Avertin (1.25% 2,2,2-Tribromoethanol, Sigma Aldrich) or with Ketamine (100mg/kg, Ketaset, Fort Dodge Animal Health) and Xylazine (10mg/kg, Rompun, Bayer Animal Health). Body length (nose to tail base) was measured and dual-energy X-ray imaging performed using a PIXImus II Bone Densitometer (GE Medical Systems, United Kingdom). The region of interest on the resulting images was manually selected to exclude the skull, and then the Lunar PIXImus software package calculated body fat mass (g), lean mass (g), fat percentage estimate (%), bone area (cm^2), bone mineral density (BMD) (g/cm^2) and bone mineral content (BMC) (g). Quality control using a phantom mouse was performed prior to imaging. The Ketamine/Xylazine anaesthesia was reversed using Atipamezole hydrochloride (1mg/kg, Antisedan, Pfizer Animal Health). To avoid introduction of bias, the experimenters were blind to the genotype during the procedure. Furthermore, cages were processed randomly and different genotypes could be housed together, hence there was no pattern to the order in which animals were processed. Mice were returned to their group housed cages and placed on a heated recovery rack until they were fully ambulant at which point they were placed on holding racks and kept for further unrelated experiments.

Datasets

To examine the relationship between body weight and the DEXA variables, wildtype data were collected for three strains, C57BL/6Dnk (B6Dnk), C57BL/6NTac (B6N) and C57BL/

6Brd-*Tyr*^*c-Brd*;C57BL/6Dnk;C57BL/6N (B6Brd;B6Dnk;B6N-*Tyr*^*c-Brd*), as detailed in Supplementary Table 1. To examine the issues in interpreting DEXA variables when body weight changes with a gene knockout, mice carrying the *Akt2*^*tm1Wcs* targeted allele were created by blastocyst injection of targeted ES cells, and bred on the 129S5/SvEvBrd genetic background. Phenotyping data was collected on 12 homozygous female mice and 17 wildtype female mice for comparison.

Data analysis

Data analysis was completed using the commercial software SPSS (Version 17.0.2) and the freeware statistical program R (23). The analysis was reported in accordance of the Animal Research: reporting of *In Vivo* Experiment (ARRIVE) guidelines (22).

Results and Discussion

Body weight was found to linearly correlate with the DEXA variables

To confirm a correlation between body weight and DEXA variables (fat mass, lean mass, fat percentage estimate, bone area, BMD and BMC), wildtype data was collected and the relationship between the DEXA variables and body weight examined with a scatter plot and a linear regression. There was a linear correlation between body weight and the DEXA variables for all three mouse strains examined (Figure. 1, Supplementary Table 2 & Supplementary Figure 1). This linear relationship had both a significant intercept and gradient element and explained on average 49% of the variation depending on the variable reviewed. The gradient and intercept estimates were similar between different wildtype strains indicating a similar relationship. Bone mineral density had the weakest correlation which was unsurprising in that as a density measure it attempts to remove the effect of animal size however there was still a correlation as bone density is known to be influenced by body weight (24). This analysis found that as body weight increased, all of the DEXA variables also increased in value and this effect was independent of genetic background. Biologically these variables were theorized to be correlated. This knowledge combined with the finding that this was seen across genetic background gives us confidence that there was a real correlation between these variables.

Comparison of statistical calls using ANCOVA and the Ratio Correction approach

Females homozygous for the thymoma viral proto-oncogene 2 (*Akt2*) targeted inactivation allele *Akt2*^*tm1Wcs*, were found to have a statistically significant decrease in both body weight and in the DEXA variables (Figure 2 & Table 1). This decrease was classified as a large effect since the Cohen's *d* measure was > 0.8 (25) for each variable (Table 1). This raised the question as to whether the changes in the DEXA variables were expected with this change in body weight (i.e. the proportion of these variables had stayed consistent), or whether the make-up of the animal had also changed alongside the body weight change.

Both ANCOVA and Ratio Correction are methodologies proposed to remove the effect of a confounding variable, thereby allowing the experimenter to assess whether the underlying variable would have changed with the treatment if the confounding variable effect was removed. Looking at the Ratio Correction approach, where the variable of interest was

divided by the body weight and a Student's *t*-test was used to compare the resulting ratios, a statistically significant difference was detected for four of the five DEXA variables assessed. Furthermore, the Cohen's *d* measure indicated that each difference was a large effect (Table 1). For lean mass, BMC and BMD, the results from the Ratio Correction approach suggested that whilst there was a drop in these DEXA variables with the knockout, once body weight was accounted for, the knockout had an increase in these variables, as the relative ratio was significantly higher in knockout mice. For fat mass, the Ratio Correction analysis suggested that once weight was accounted for, there was a decrease in the amount of fat mass per unit weight in knockout mice. In contrast, when the effect of body weight on the DEXA variables was removed using the ANCOVA method, no statistically significant difference was detected for any of the DEXA variables (Table 1). In summary, whilst targeted inactivation of *Akt2* clearly affected body weight, the Ratio Correction method suggested that the knockout also had effects on the DEXA variables to body weight proportions, whilst ANCOVA analysis indicated that there was no significant difference in the DEXA variables once the relationship with body weight was accounted for.

Are the results from the Ratio Correction approach meaningful?

As with all statistical tests, the Ratio Correction approach makes a number of assumptions that affect the performance of the test. Specifically, it assumes a linear relationship with the confounding variable through the origin (i.e. the relationship can be described with a gradient term and not a significant intercept term), that the ratio is normally distributed, that the measurements are independent, and finally that the variance in the two groups being compared is equivalent (homogeneity of variance). From wildtype data, a linear relationship was demonstrated such that as body weight increased so did the DEXA variables (Figure 1, Supplementary Table 2 & Supplementary Figure 1). This relationship can be described by a linear equation which has both a significant intercept and gradient element (Supplementary Table 2) (Eq[1].), where x is the body weight, y is the DEXA variable of interest, c is the intercept, and m the gradient.

$$y = mx + c \quad \text{[Eq. 1]}$$

In the Ratio Correction approach, the variable of interest (y) is divided by the confounding variable (x). If the linear relationship can be described with only a significant gradient, then on division by x, the ratio becomes equal to the gradient (Eq[2].).

$$\frac{y}{x} = \frac{m\cancel{x}}{\cancel{x}}$$

[Eq. 2]

When the linear relationship is described by both a significant gradient and an intercept then the dependency on x is not removed. Rather the ratio becomes inversely dependent on x, such that as the confounding variable increases the ratio decreases (Eq[3].). This happens as

the confounding variable is now dividing the intercept component, so as it increases, the element added to the gradient element gets smaller, and so the calculated ratio decreases.

$$\frac{y}{x} = m + \frac{c}{x} \quad \text{[Eq. 3]}$$

Therefore in the comparison of wildtype and *Akt2* knockout mice, where the relationship between weight and the DEXA variable had both a significant intercept and gradient, the calculated ratio decreased significantly between the wildtype and the knockout mice, and a negative Cohen's *d* value was seen as the dependency on body weight had not been removed. There were two exceptions to this. Firstly, the Ratio Correction approach on the fat mass variable gave a positive Cohen's *d*, indicating the calculated ratio increased from the wildtype to the knockout mice. This was explained by looking in more detail at the linear relationship between fat mass and body weight in the *Akt2* dataset (Eq[4].). Since the intercept element (−6.7) was negative, and the body weight was dividing the intercept value, as the body weight increased the amount being subtracted decreased, and hence the calculated ratio increased (Eq[5].). The key component in this instance was the negative intercept, which gave rise to the positive Cohen's *d* value that was observed.

$$y = 0.6x - 6.7 \quad \text{[Eq. 4]}$$

$$\frac{y}{x} = 0.6 - \frac{6.7}{x} \quad \text{[Eq .5]}$$

The second exception was that fat percentage estimate was not significant after applying the Ratio Correction approach, but was significant when comparing wildtype and knockout raw, unadjusted data (Table 1), indicating that the effect of the confounding variable had been removed by the Ratio Correction approach. This was explained by looking in more detail at the linear relationship between fat percentage estimate and body weight in the *Akt2* dataset (Eq[6].).

$$y = 1.3x - 0.14 \quad \text{[Eq. 6]}$$

$$\frac{y}{x} = 1.3 - \frac{0.14}{x} \quad \text{[Eq. 7]}$$

For this variable, the dominant component in the relationship was the gradient (1.3), whilst the intercept component, which gave a dependency on body weight even following the normalization step, was not a statistically significant component of the relationship, contributing $< 1\%$ of the total signal seen. Therefore a change in body weight between the wildtype and knockout mice did not have a statistically significant impact on the ratio, and thus the Ratio Correction approach, as per equation 2, removed the dependency on body weight. The key component in this instance was the negligible intercept value (−0.14).

A basic assumption of the Ratio Correction approach is that a linear relationship through the origin exists between the variable of interest and the confounding variable. The *Akt2* data clearly provides an example where the Ratio Correction approach can generate misleading results when this assumption is not met.

Are the results from the ANCOVA approach robust?

ANCOVA uses regression analysis to adjust for the confounding variable by first estimating the linear relationship between the confounding variable and the variable of interest, and then testing whether the adjusted group means vary significantly from each other. The ANCOVA assumes homogeneity of variance, normality, independent sampling of the measurements, and a linear relationship between the confounding variable and the variable of interest. This is similar to the Ratio Correction approach, except that these assumptions apply to the raw data and the linear relationship can have both a significant intercept and gradient. The ANCOVA has two additional assumptions that are not required for the Ratio Correction approach. These are homogeneity of regression slope, and independence of the confounding variable and the treatment.

The linearity and homogeneity of regression slope assumptions were assessed together as follows. Scatterplots of the data (Supplementary Figure 1) supported the assumption that, for wildtype mice, a linear relationship existed between body weight and the DEXA variables. This assumption was tested for the *Akt2* dataset by visually plotting the data (Supplementary Table 3) and by examining whether the confounding variable was statistically significant ($p<0.05$) in the ANCOVA analysis (Supplementary Table 4). The homogeneity of regression slope assumption arises because the ANCOVA approach fits a regression line to the entire dataset, ignoring which group the data point belongs to, and thus it is assuming that this linear relationship is true for all groups. When the numbers of readings in each group are equal, the ANCOVA is considered to be a sufficiently robust method to withstand violation of the homogeneity of regression assumption (26). If, however, there is gross violation of this assumption, then this should be investigated as this in itself is biologically interesting. This assumption can be assessed by plotting the confounding variable (body weight) against the outcome (phenotypic measure) and comparing the gradients on regression lines fitted for each experimental condition (genotype) (Fig. 3). The assumption can also be tested by customizing the ANCOVA test to add an interaction term between the confounding variable and the treatment variable and assessing whether the interaction is significant (see Figure 3 for an example and full details in Supplementary Table 3). Testing for an interaction term found that the assumption of homogeneity of regression slope was met with the *Akt2* dataset. Visual examination of the data found that the regression lines were very similar however they will intersect at higher body weights (Fig. 3). There is low power to detect a difference in regression slope, but in our opinion the visual difference in slope is not biologically significant and could be arising from a sampling issue or due to a minor difference in the regression slope.

The linear relationship underpinning the ANCOVA, includes an intercept element that biologically is not plausible. For example, in our scenario a mouse of zero body weight is not going to have minimum body composition. The poorly fitting regression at low signal

arises as many morphological and physiological variables are scaled relative to the body size, according to a power equation (see Eq[8]. for general form). This is described as allometric scaling (27). The power equation can be converted to a linear form by a log transformation of Y and X (Eq[9]). In a situation of allometric scaling, where Eq.8 represent the relationship, an ANCOVA will be more precise when the dependent variable and body weight variable are logged; as the underpinning relationship would have been modeling more accurately (28). This approach has a number of assumptions. Firstly, that animals are geometrically similar and secondly that Eq[9.] defines the relationship. As the regression assumption was met in this analysis over the dynamic range the body weight varied, an allometric approach was not applied.

$$Y = a.X^{b} \quad \text{[Eq. 8]}$$

$$Log\,Y = log\,a + b\,log\,X \quad \text{[Eq. 9]}$$

The second ANCOVA specific assumption, that the confounding variable and treatment effects are independent, can be met provided that the experimental conditions do not affect distribution of the confounding scores, and that the confounding variable does not influence the nature of the experimental condition. This assumption is frequently violated in instances where ANCOVA is used specifically to "equate" groups that differ on a confounding variable rather than the more common use of reducing the variance. In the present study, this assumption was clearly violated as shown by a statistically significant difference between the knockout and control mice for body weight, the confounding variable concerned (Table 1). Trying to remove the effect of a confounding variable is risky, because in the process you may remove part of the treatment effect (19, 26, 29). Wright argued that ANCOVA does have a role in situations where a confounding variable varies with the treatment but highlights the need to be clear about which research question is being asked and which assumptions are being made (1). Mathematically the ANCOVA is assuming a common regression line and when groups are distinct this line is being extrapolated from one group to the other. Therefore there is a risk that the line fitted is one that removes the treatment effect rather than the underlying concerning correlation between the two variables. This effect, unlike the issues with the ratio normalization, is a conservative effect as it leads to false negatives rather than false positives. To examine the risk that the regression slope being fitted within the ANCOVA is removing more than the confounding relationship, visual inspection of the data is therefore essential to ensure that the overall linear fit would be appropriate for each group independently, and only then is it valid to ignore this assumption and complete the analysis cautiously. This was completed for the *Akt2* dataset by examining the scatterplots (Figure 3 & Supplementary Table 3) which confirmed that a linear relationship existed between the confounding and measured variable that was not dependent on the genotype. The ANCOVA, therefore, was robust to the violation of the assumption that the treatment and confounding effects were independent.

As the ANCOVA specific assumptions were met for the *Akt2* dataset, the next step was to assess the more general assumptions of independence, normality and homogeneity of variance. The assumption of independent sampling was met as the experiment was designed

such that one reading was taken from one mouse and we know of no significant reasons to expect clusters in the data. The final assumptions of homogeneity of variance and normality were assessed using graphical tools to inspect the distribution of standardized residuals. Standardized residuals are the difference between observed and expected values from the fitted ANCOVA after they have been standardized for the variability in the data (Figure 4 & Supplementary Table 5). No concerning bias in residuals was seen across the treatment groups, suggesting the model was a good fit and the assumptions were appropriate. Since the assumptions required for a valid ANCOVA were met in the *Akt2* dataset, the outcome should be robust and, consequently, accepted.

Conclusions

The presence of confounding variables can lead to an association that, whilst true, is potentially misleading. This is commonly seen during phenotypic characterization of genetically altered lines of mice, such as knockouts, where multiple abnormalities can be detected and the experimenter will want to assess whether there is a significant difference once the relationship with the confounding variable was accounted for. The Ratio Correction approach, often referred to as Normalization, is a highly risky approach that is commonly used. The problem with this approach is that the underlying assumptions become hidden and, when these assumptions are not met, the outcome lacks meaning and can lead to inappropriate biological conclusions. This was seen in the *Akt2* knockout dataset where the Ratio Correction approach suggested that, after adjusting for weight, an increase in some of the DEXA variables was biologically significant, thus leading to the conclusion that the knockout had a direct impact on the body composition along with a change in body weight. In contrast, ANCOVA suggested that the changes in the DEXA variables were all associated with the change in body weight seen in this knockout line.

Kilkenny and coworkers in a review of animal studies, found that testing the assumptions of the analysis methods used was rarely completed (30). The authors argued that it was a requirement for robust biological findings and consequently the ARRIVE guidelines include this as a required component for publication to ensure compliant with the 3R's (22). This manuscript demonstrates how assumption testing is essential to ensure an analysis method is robust and therefore the results reliable. In this case, both ANCOVA and Ratio Correction assume a linear relationship between the confounding variable and variable of interest. However, the Ratio Correction method commonly fails because it assumes a linear relationship through the origin. So whilst the Ratio Correction method is simplistic to apply, the outcome can be misleading and, in our opinion, this methodology should not be used. Instead, ANCOVA with assumption testing is the process by which researchers should be assessing this type of data. The sample size frequently used in animal experiments is often low and can reduce the power of ANCOVA, however this is also an issue for any other statistical approach and hence this is not an argument against using the ANCOVA. As a first step, the researcher needs to test the basic premise i.e. assess whether a linear relationship actually exists between the confounding variable and the variable of interest. An example of this was raised by Yin *et al.* who discussed that body weight was thought to be a confounding variable for heart weight, as biologically larger animals would be expected to have large hearts. However, frequently no linear relationship could be seen between body

weight and heart size, and instead they proposed tibia length as a more appropriate confounding variable (4). ANCOVA at times is not an ideal method as it assumes a single linear relationship exists and in situations where the groups are quite different, as in this phenotyping example, there are risks with this assumption and the accuracy of this calculation. However, as this effect is conservative and no other technique is available, we support the use of ANCOVA to address confounding factors within datasets. Unlike Ratio Correction, ANCOVA relies on the use of statistical packages. InVivoStat, a free-to-use R based statistical software package aimed at animal researchers includes ANCOVA as a tool (31) and automatically generates the diagnostic plots along with text explaining what the user should look for and how the assumptions can be assessed.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Wright DB. Comparing groups in a before-after design: When t test and ANCOVA produce different results. Brit J Educ Psychol. 2006; 76:663–75. [PubMed: 16953968]

2. Brown SD, Hancock JM, Gates H. Understanding mammalian genetic systems: the challenge of phenotyping in the mouse. PLoS genetics. 2006; 2(8):e118. [PubMed: 16933996]

3. Reed DR, Lawler MP, Tordoff MG. Reduced body weight is a common effect of gene knockout in mice. BMC genetics. 2008; 9:4. [PubMed: 18182103]

4. Yin FC, Spurgeon HA, Rakusan K, Weisfeldt ML, Lakatta EG. Use of tibial length to quantify cardiac hypertrophy: application in the aging rat. The American journal of physiology. 1982; 243(6):H941–7. [PubMed: 6216817]

5. Whitnall M, Survo Rahmanto Y, Sutak R, Xu X, Becker EM, Mikhael MR, et al. The MCK mouse heart model of Friedreich's ataxia: Alterations in iron-regulated proteins and cardiac hypertrophy are limited by iron chelation. Proceedings of the National Academy of Sciences of the United States of America. 2008; 105(28):9757–62. [PubMed: 18621680]

6. Sun X, Zhou Z, Kang YJ. Attenuation of doxorubicin chronic toxicity in metallothionein-overexpressing transgenic mouse heart. Cancer research. 2001; 61(8):3382–7. [PubMed: 11309296]

7. Ryu KY, Garza JC, Lu XY, Barsh GS, Kopito RR. Hypothalamic neurodegeneration and adult-onset obesity in mice lacking the Ubb polyubiquitin gene. Proceedings of the National Academy of Sciences of the United States of America. 2008; 105(10):4016–21. [PubMed: 18299572]

8. Kashimada K, Yamashita T, Tsuji K, Nifuji A, Mizutani S, Nabeshima Y, et al. Defects in growth and bone metabolism in klotho mutant mice are resistant to GH treatment. The Journal of endocrinology. 2002; 174(3):403–10. [PubMed: 12208660]

9. Melgar S, Bjursell M, Gerdin AK, Svensson L, Michaelsson E, Bohlooly-Y M. Mice with experimental colitis show an altered metabolism with decreased metabolic rate. Am J Physiol Gastrointest Liver Physiol. 2007; 292(1):G165–72. [PubMed: 16844678]

10. Liermann M, Steel A, Rosing M, Guttorp P. Random denominators and the analysis of ratio data. Environmental and ecological statistics. 2004; 11:55–71.

11. Nevill AM, Holder RL. Scaling, normalizing, and per ratio standards: an allometric modeling approach. Journal of applied physiology. 1995; 79(3):1027–31. [PubMed: 8567498]

12. Albrecht G. Some comments on the use of ratios. Syst Zool. 1978; 27(1):67–71.

13. Tanner JM. Fallacy of per-weight and per-surface area standards, and their relation to spurious correlation. Journal of applied physiology. 1949; 2(1):1–15. [PubMed: 18133122]

14. Reist J. An empirical evaluation of several univariate methods that adjust for size variation in morphometric data. Can J Zool. 1985; 63:1429–39.

15. Stevens MT. Value of Relative Organ Weights. Toxicology. 1976; 5(3):311–8. [PubMed: 944470]

16. Shirley E. Analysis of Organ Weight Data. Toxicology. 1977; 8(1):13–22. [PubMed: 929613]

17. Goran MI, Allison DB, Poehlman ET. Issues relating to normalization of body fat content in men and women. Int J Obes Relat Metab Disord. 1995; 19(9):638–43. [PubMed: 8574274]

18. Poehlman ET, Toth MJ. Mathematical ratios lead to spurious conclusions regarding age- and sex-related differences in resting metabolic rate. The American journal of clinical nutrition. 1995; 61(3):482–5. [PubMed: 7872210]

19. Fields, A. Discovering statistics using SPSS. 3rd ed.. SAGE Publication Ltd; London: 2009. p. 822

20. Okura T, Tanaka K, Nakanishi T, Lee DJ, Nakata Y, Wee SW, et al. Effects of obesity phenotype on coronary heart disease risk factors in response to weight loss. Obesity research. 2002; 10(8):757–66. [PubMed: 12181384]

21. Tschop MH, Speakman JR, Arch JR, Auwerx J, Bruning JC, Chan L, et al. A guide to analysis of mouse energy metabolism. Nature methods. 2011; 9(1):57–63. Epub 2011/12/30. [PubMed: 22205519]

22. Kilkenny C, Browne WJ, Cuthill IC, Emerson M, Altman DG. Improving Bioscience Research Reporting: The ARRIVE Guidelines for Reporting Animal Research. Plos Biol. 2010; 8(6)

23. Ihaka R, Gentleman R. R: A Language for Data Analysis and Graphics. Journal of Computational and Graphical Statistics. 1996; 5(3):299–314.

24. Felson DT, Zhang YQ, Hannan MT, Anderson JJ. Effects of Weight and Body-Mass Index on Bone-Mineral Density in Men and Women - the Framingham Study. Journal of Bone and Mineral Research. 1993; 8(5):567–73. [PubMed: 8511983]

25. Cohen, J. Statistical power analysis for the behaviour sciences: Lawrence Earlbaum Associates. 1988.

26. Owen SV, Froman RD. Uses and abuses of the analysis of covariance. Research in nursing and health. 1998; 21(6):557–62. [PubMed: 9839800]

27. Schmidt-Nielsen, K. Scaling - Why is animal size so important?. Cambridge University Press; Cambridge: 1985. p. 245

28. Nevill AM, Ramsbottom R, Williams C. Scaling Physiological Measurements for Individuals of Different Body Size. Eur J Appl Physiol Occup Physiol. 1992; 65(2):110–7. [PubMed: 1396632]

29. Wu YW, Slakter MJ. Analysis of covariance in nursing research. Nursing research. 1989; 38(5):306–8. [PubMed: 2798159]

30. Kilkenny C, Parsons N, Kadyszewski E, Festing MF, Cuthill IC, Fry D, et al. Survey of the quality of experimental design, statistical analysis and reporting of research using animals. PloS one. 2009; 4(11):e7824. [PubMed: 19956596]

31. Clark RA, Shoaib M, Hewitt KN, Stanford SC, Bate ST. A comparison of InVivoStat with other statistical software packages for analysis of data generated from animal experiments. J Psychopharmacol. 2011 Epub 2011/11/11.
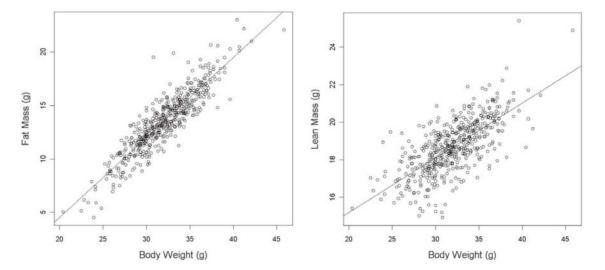
Figure 1. Representative scatter plots to show the correlation between body weight and the DEXA variables.

Data shown are from B6Brd;B6Dnk;B6N-*Tyr*^{c-Brd} wildtype female mice for fat mass (A)
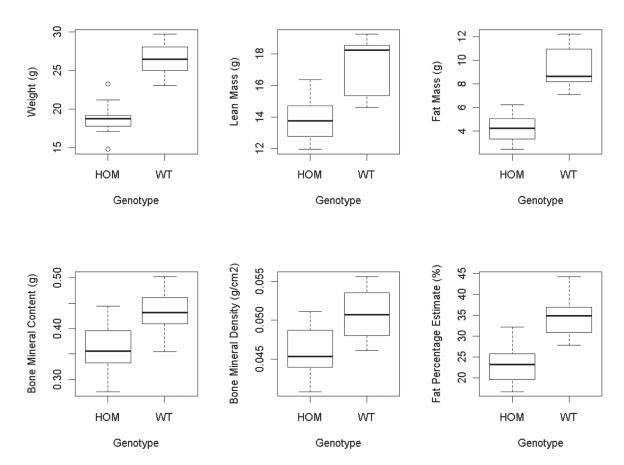
and lean mass (B).

Figure 2. Boxplot comparisons of the impact of *Akt2* genotype (wildtype (WT) controls versus *Akt2^tm1Wcs* homozygotes (HOM)) on body weight and DEXA variables.

The whiskers extend to the most extreme data point which is no more than 1.5 times the inter-quartile range. Points beyond this are classed as outliers and are shown individually as open circles.
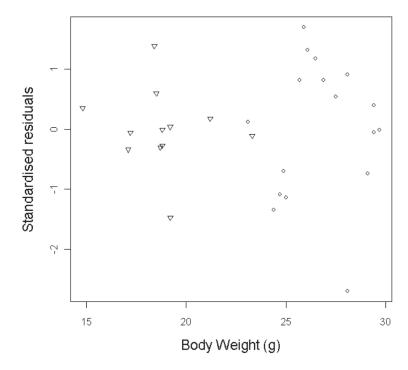
Figure 3. Testing the assumption of homogeneity of variance.
An example scatterplot, from *Akt2[tm1Wcs]* homozygous mice versus control study, of the confounding variable (body weight) against the DEXA variable of interest (lean mass) which has a regression line fitted for each treatment group (control shown with triangles versus knockout shown with circles) visually suggests that the regression line is equivalent between the two groups. An ANCOVA which included an interaction term calculated a *p*-value for this effect of 0.702, meaning this term would not be considered a significant component of the data and that the slopes are equivalent between the two groups. The ANCOVA calculated a *p*-value for the body weight as a confounding factor as $p < 0.000$ which indicates it is a highly significant component of the analysis.
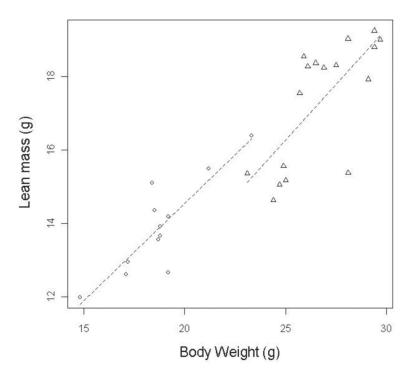
Figure 4. Representative standardized residuals scatterplot for the *Akt2^{tm1Wcs}* homozygous mice versus control dataset.

In this example, the standardized residuals for the lean mass variable were plotted against body weight. Data from knockout mice are shown with a triangle and controls with a circle.

Table 1

Statistical significance when assessed with different approaches

The statistical outcome for the comparison of body weight and DEXA variables for female *Akt2^tm1Wcs* mice is presented when the data was analyzed using classic statistical tests with raw and Ratio Corrected data, and then using ANCOVA. Cohen's *d* is a measure of the strength of the treatment effect and is calculated as the difference between the means (control mean – knockout mean) divided by the average standard deviation (25). A measure of the treatment effect was only calculated in situations where a statistically significant difference was detected.

Variable	Raw data analysis (Student's *t*-test)		Ratio Correction (Student's *t*-test on ratio)		ANCOVA analysis of raw data
	p-value	Cohen's *d*	*p*-value	Cohen's *d*	Genotype effect: *p*-value
Body weight	0.000	3.9	-	-	-
Lean Mass	0.000	2.3	0.000	−2.2	0.188
Fat Mass	0.000	3.4	0.000	2.5	0.450
Bone Mineral Content	0.000	1.6	0.000	−1.6	0.887
Bone Mineral Density	0.000	1.5	0.000	−3.0	0.750
Fat percentage estimate	0.000	2.5	0.508	-	0.228