Accepted manuscript

# The development and evaluation of multiple regression equations based on four common nutritional analysis packages to predict the metabolisable energy density of diets fed to grower/finisher and adult pigs and their use for rat and mouse diets

Graham Tobin[1], Annette Schuhmacher[2], Tomasz Górecki[3], Łukasz Smaga[3]

[1]The Orchard, Weeping Cross, Bodicote, United Kingdom, OX15 4EE
[2]ssniff Spezialdiaeten, Ferdinand-Gabriel Weg 16, 59494 Soest, Germany
[3]Faculty of Mathematics and Computer Science, Adam Mickiewicz University, Uniwersytetu Poznańskiego 4, 61-614 Poznań, Poland

**Corresponding Author:** Dr Graham Tobin, The Orchard, Weeping Cross, Bodicote, United Kingdom, OX15 4EE, gtobin500@gmail.com, +44 1295 262064

**Keywords**: Metabolisable energy density prediction, Multiple linear regression analysis, Pigs Rats and mice


This peer-reviewed article has been accepted for publication but not yet copyedited or typeset, and so may be subject to change during the production process. The article is considered published and may be cited using its DOI
10.1017/S0007114525000042
The British Journal of Nutrition is published by Cambridge University Press on behalf of The Nutrition Society

## Abstract

We have used multiple regression analyses to develop a series of metabolisable energy (ME) prediction equations from chemical analyses of pig diets that can be extended to murine diets. We compiled four datasets from an extensive range of published metabolism studies with grower/finisher and adult pigs. The analytes in the datasets were increasingly complex, comprising: 1. the proximate or Weende analysis, 2. the previous analysis but with neutral detergent fibre (NDF) replacing crude fibre, 3. the NDF package plus starch, 4. the NDF package plus starch and sugars. Diet manufacturers routinely provide most of the analytes for batches of murine diet, or they are easily obtainable. The study uniquely compares the four analytical packages side-by-side. The number of records in the datasets varies from 367 to 827. With increasing analytical complexity, adjusted $R^2$ values for ME prediction improved from 0.751 to 0.869, and the mean absolute error from 0.422 to 0.289 kJ/g. Overall, the models' prediction interval (PI) improved from 1 to 0.7 kJ/g, which is ± 7% to 5% for a typical dietary ME density of 14.8 kJ/g. Although prediction accuracy increases as one extends the range and complexity of the analytes measured, the improvement is slight and may not justify the substantial increase in analytical cost. The equations were validated for use on future data sets by k-fold analysis. Although the equations are developed from pig data, they are suitable for rat and mouse diets, based on comparable digestibility measurements, and substantially improve existing methods.


## Abbreviations

CP - Crude Protein, EE - Ether Extract, AEE- Acid Hydrolysis Ether Extract, CF - Crude Fibre, NFE - Nitrogen-free Extractives, NDF - Neutral Detergent Fibre, NFC - Non-fibrous Carbohydrates, Res1 - Residue 1, Res2 - Residue 2, GE - Gross energy, DE - Digestible Energy, ME - Metabolisable Energy. The residues are the calculated difference between the total dry matter and measured nutrients.

## Introduction

Many murine research studies report digestible energy (DE) and/or metabolisable energy (ME) intake of natural-ingredient diets. These should preferably be determined in an *in vivo* study from food intake, faecal output, and, for ME intake, urine output, and each component's gross energy (GE) density [1]. These measurements require specialist equipment and cages and appropriate expertise [2; 3; 4; 5; 6]. Only a few laboratories that regularly perform energy metabolism studies meet these requirements. Consequently, investigators often estimate energy intake by one of two indirect methods.

First, they may multiply the food intake by the diet's energy density estimated from Atwater Factors applied to the composition of the diet [7]. The values are 4 kcal/g protein and carbohydrate (carbohydrate includes fibre) and 9 kcal/g fat; the equivalent SI values are about 16.7 and 37.7 kJ/g, respectively. Atwater obtained the values (also called Physiological Fuel Values) from studies with humans fed typical diets. Subsequently, Merrill and Watt [8] slightly modified the values for different foods. Despite the widespread use of Atwater Factors, there are doubts about their accuracy, even for human foods [8; 9; 10]. Furthermore, they assume a high nutrient digestibility, typical of refined human foods. Bielohuby and colleagues [11] have confirmed that Atwater Factors are inappropriate for lower digestible, natural-ingredient murine diets. Second, investigators may also use ME density estimates from diet manufacturers' data sheets. However, with a few exceptions, these are based on Atwater Factors and have the same disadvantages.

At about the same time as Atwater's work on human foods, agriculturists were also developing a means of predicting the energy content of natural-ingredient animal feed and diets that allowed for differences in digestibility. Foremost in these methods was estimating ME density from total digestible nutrients (TDN), in which Atwater Factors were fitted to the digestible nutrients [12; 13]. However, since the 1960s, direct measurement of DE and ME has become the standard method of energy evaluation in the United States and most of Europe [14]. This change in energy evaluation methodology and the widespread availability of computers in the 1960s led to the development of energy predictions from multiple regression analysis, with dietary nutrients as covariates. Initially, the covariates were digestible proximate or Weende nutrients, usually to comply with regulatory requirements [15]. The Rostock group was at the forefront of developing such predictions in the 1960s. Their data has been summarised in a compendium by Schiemann [16] and two recent reviews [17; 18]. Their work formed the

basis of the currently recommended means of predicting the ME density of pig diets in Germany from digestible [19] and crude nutrients [20].

Routine measurement of digestible nutrients is impracticable and arguably more difficult than directly determining ME density. Consequently, chemical analysis of crude nutrients began to replace digestible nutrients in prediction equations [21]. The crude nutrients ranged in complexity from those in the simple proximate or Weende analysis (crude protein, ether extract, crude fibre, ash, and NFE) to those based on more complex fibre types (NDF, ADF) and soluble carbohydrates (starch, sugars) [15; 22; 23; 24; 25]. While each laboratory provides one or more valid equations, they reflect the laboratory's specific conditions, such as the age and weight of animals, the techniques used, and environmental conditions. Applying predictions more widely from an individual laboratory underestimates the likely prediction errors [24]. Unfortunately, investigators often fail to consider the size of prediction errors when drawing conclusions from studies in which one or more variables, such as energy intake, are estimated rather than measured.

Our study aimed to improve ME density estimates of murine diets. We had hoped to base these estimates on data obtained in rats and mice, but this proved impractical. Consequently, the emphasis switched to the suitability of data from pig studies for murine prediction equations. Several studies suggested this might be practicable [26; 27; 28].

Our first step was to confirm the similarity of digestibility of nutrients in the pig, rats, and mice that would justify using predictions of DE and ME density of natural-ingredient diets from pig studies for murine models. Second, if so, we intended to create prediction equations for energy density in pigs from many published studies using grower/finisher and adult pigs. The predictors were to be those commonly reported in four typical analytical packages for diets. The number of data records is greater than in previous studies. All the data are measured, not calculated, from various geographical areas and research groups. Two other recent publications have taken a similar approach by obtaining data from a wide range of sources [29; 30], though there were no equivalent prediction equations for dietary ME density. A critical function of our study was to estimate how well our equations fitted the existing data and their accuracy in predicting dietary ME density from future analytical data.

## Materials and Methods

The study comprised two phases, each requiring distinct sets of data. These phases are listed below, along with a description of the data used and statistical analyses applied to the data. Typically, we used R Statistical Software v 4.2.3 [31] and its various packages. We installed the

packages and loaded their library of procedures into RStudio (http://www.rstudio.com/). We provide the list of packages in the Supplementary Material. We occasionally supplemented these methods with statistical software packages described in the text. Although R is open source software, we recognise that some readers prefer 'plug and play' software to using such code. We identify comparable software in the Supplementary Material. Some of our methods are not readily available, and we have provided the code we used in those cases.

***Phase 1. Evidence that energy digestibility in pigs and rats is similar.***

Energy digestibility is the major contributor to metabolisable energy (ME) density. To justify using pigs as a model for murine ME density, we must first demonstrate good agreement between the DE or ME density of diets fed to murine and pig models. We obtained 12 papers containing measurements of energy digestibility in diets or ingredients fed to both pigs and rats. These papers contributed 204 data pairs (we excluded two pairs, based on feeding palm kernel products, as substantial outliers). The references are given in the Supplementary Material. We used the Bland-Altman Plot to assess the agreement in energy digestibility between pigs and rats [32; 33; 34].

***Phase 2. The development of predictive estimates of dietary ME density based on dietary chemical analysis.***

a. Data sources and selection

We used PubMed, Google Scholar, and apt journals to obtain papers that included measurements of ME density and relevant chemical analyses. The list of papers used is provided in the Supplementary Material. For consistency, we refer to each entry containing the analytical observations in a diet or ingredient as a record and the collection of records as a database. We describe any subgroup of the database as a dataset. The first division of the database is into four datasets representing different chemical analysis groups. For simplicity, we refer to the groups in the text as the CF, NDF, NDFS, and NDFSS datasets. These four datasets include measurements of dry matter, crude protein, ether extract, ash and one of crude fibre, neutral detergent fibre, neutral detergent fibre plus starch, or neutral detergent fibre plus starch and sugars. Table 1 shows the nutrient components of the four groups. Each of the four groups includes a residue value, the difference between the total dry matter and the sum of the measured analytes. The residues were nitrogen-free extractives, non-fibrous carbohydrates, and Residues 1 and 2. The number of papers used in each dataset was: CF, 55; NDF, 80; NDFS, 31; and NDFSS, 8. Several papers contained all these carbohydrate

measurements. The data were standardised to a 100% dry matter basis and expressed as g/100 g for chemical analytes and kJ/g for energy density.

With a few exceptions, we excluded ingredients from our source data. Although several investigators have argued that chemical analysis and energy density measurements of ingredients can contribute to dietary energy density predictions (additivity) [21; 23; 35; 36; 37; 38; 39], it is not a universal view. Additivity may be invalid in several circumstances. First, when ingredients have a nutritional profile very different from the mixed diet [40] or when mutual supplementation of proteins or amino acids supplementation affects urinary energy excretion and metabolisability [41; 42]. Finally, the methods used to determine the energy density of ingredients may commonly contain errors [43; 44; 45]. We had sufficient records from diets to generally avoid the need to include ingredient data. The exceptions were those where the ingredient contributed to almost the entire diet, only supplemented with minerals and vitamins.

Murine diets.

We found 34 records in six papers that reported measured dietary ME density and appropriate chemical analysis obtained with rats or mice, though for the CF model only. The number of records is too few to provide reliable regression equations: a minimum of 10-20 records is required per independent variable [46] and possibly more [47]. Consequently, we limited our data to those from pig studies.

Pig diets

We only included data from pigs in the grower/finisher (25 to 125 kg) and adult phases (usually sows weighing over 200 kg). For most analyses, we combined the data from the two groups. Occasionally, authors duplicated records in more than one paper, and we excluded them to prevent any bias or data weighting. We also eliminated outliers using the methods described in statistics. Regrettably, we excluded many potentially valuable papers because they lacked some analytical component(s), often ether extract or ash. Sometimes, we could retain such papers with additional information from the authors.

b. Statistical analyses

*Data description*

We used the R datasummary function to obtain descriptive statistics of the four analytical datasets for the combined grower/finisher (GF) and adult (AD) phases. These descriptives are shown in Table 1.

*Analysis of variance (ANOVA)*

Where an ANOVA was necessary, we first tested for normality using a Q-Q plot and the Shapiro-Wilks test. The appropriate ANOVA test depended on the outcome, with the Welch ANOVA test preferred for parametric data and the Kruskal-Wallis test for nonparametric data.

*Removal of outliers with the isolation forest method and extreme standardised residuals*

We excluded outliers from the four datasets using the isolation forest (IF) method [48] as described here. We initially used multiple regressions to analyse the records in each dataset (CF, NDF, NDFS, and NDFSS). The dependent variable was ME density, with the chemical analyses as the covariates. The analysis created a predicted ME density and its standardised residual (the distance of the actual value from the prediction divided by the standard deviation of the residuals) for each observation. We then applied the isolation forest procedure to the pairs of predicted ME density and standardised residuals using the solitude package in R. The isolation forest process can be envisaged as randomly drawing horizontal and vertical lines through a scatter plot. The number of lines required to isolate each point from adjacent ones is recorded. The exercise is repeated, in our case, 100 times, but 'starting' the line in a different location each time. Only a few lines are necessary to separate outliers, while those in a cluster need many more (see the illustration in Liu *et al.* [48]). The process then allocates an anomaly score to each point based on the average number of 'cuts' or divisions required to isolate it. The score is standardised and ranges from zero (not anomalous) to one (anomalous). The dividing score between points being definitely anomalous or not anomalous is 0.5. We considered records with an anomaly score of about 0.65 and above as outliers. We used a rigid procedure to ensure an unbiased exclusion of outliers across all datasets.

*Multiple Regression Analysis*

We used several R packages and functions to obtain the statistics for the multiple regressions and goodness of fit measurements. We have included a list and their purpose in the Supplementary Material.

Collinearity and variable selection

Collinearity occurs when a regression has a high correlation between two or more predictors [49]. It is generally assumed to increase the confidence interval and p-value and decrease the precision of regression coefficients. Compositional data such as ours (where the predictors add up to 1 or 100) presents a severe problem with collinearity, and regression analysis using all the predictors becomes impracticable. We applied the standard solution of removing a

predictor - the ‘drop one’ approach. This process, including the choice of predictor, is discussed further in the correlation matrices section of the results.

We assessed the risk of collinearity in two ways. First, we used a correlation matrix to examine the degree of correlation between pairs of predictors. Although we describe the pair-wise correlations in the results, the tables are in the Supplementary Material. If there is significant collinearity, the highest correlated pair usually becomes the candidate from which one predictor is dropped. Secondly, we determined each coefficient’s variance inflation factor (VIF), which is a more effective method. The VIF reports the overall association between predictors rather than between pairs, as shown in the correlation matrix. One would typically consider dropping the predictor with the highest VIF to avoid collinearity. We tested for detrimental effects of collinearity on coefficients using a measure of SE adjusted for their size. We term this the coefficient of variation of the SE (CVse), calculated as 100xSE/mean. The values are shown in the Supplementary Material.

Assessment of potential laboratory bias

To test for potential bias in our regressions from data from one or more laboratories, we allocated each record to one of eight Lab groups that might use similar pigs and have a common approach to environmental conditions and techniques, e.g. Noblet’s group. Lab group 0 was the reference group made up of a large number of records from unconnected research groups. We measured potential laboratory bias from the variation in specific laboratory EMM and contrasts after multiple regression analysis with Laboratory as a factor. The estimated marginal mean (EMM) is the mean ME density when one applies the regression equation to the subset of predictors for a particular laboratory. The factor also gives the difference (contrast) between the EMM value for each laboratory and the reference Lab 0 (see Supplementary Material). We statistically compared the-EMMs and contrasts before and after outlier removal across the four datasets.

Regression analysis methods using intercept and no-intercept models

After checking for bias and outlier removal, we performed intercept and no-intercept (regression through the origin, RTO) multiple linear regressions for the four analytical datasets for both GE and ME density. Regression models are rarely used to predict GE density. However, the model is a valuable check on the data and regression model since one can match the predicted coefficients against well-established estimates of the energy content of the primary nutrients.

Intercept model: We analysed the combined GF and AD data incorporating Phase (GF, AD) as a factor. The regression analysis gives identical nutrient coefficients for both grower/finisher and adult animals. The coefficient for the Phase represents the additional energy adult animals obtain from a diet compared to the grower/finisher animals. The additional energy should only be added to ME density predictions for adult animals.

No intercept model: Many statisticians advise against the no-intercept model since it weakens the regression fit to the data [50]. However, we use it here to estimate the relative contribution of energy-containing variables to the overall energy density [51; 52], and thus ash is excluded as an independent variable. Unfortunately, the no-intercept model cannot account for external variables, such as the growth phase, effectively substituting these variables with a pseudo-intercept value. Consequently, when we used a no-intercept model for ME density, the grower/finisher and adult animal data were analysed separately. That was not necessary for GE density since this is unaffected by Phase.

Goodness of fit estimates

We determined various measures of goodness of fit (GOF) of the regressions. We used the adjusted coefficient of determinations ($R^2$, the Ezekiel estimator [53]) to express the proportion of variation in the dependent variable accounted for by the independent variables. Unlike the coefficient of determination ($R^2$) derived from the Pearson correlation coefficient (r), it compensates for what otherwise would be an increase in value caused simply by adding additional independent variables. Adjusted $R^2$ thus allows us to compare the GOF in models with different numbers of independent variables. We also included predicted $R^2$ (see the section on validation). However, we excluded R values for the no-intercept models since they are inflated by the standard calculation method and provide an unreliable estimate of their relative performance [54].

We also determined the root mean square error (RMSE) and residual standard deviation/error (RSD/RSE/sigma are used interchangeably). Although RMSE and RSD are closely related, they have slightly different uses. Outliers affect RMSE more than RSD, and a few extreme values may introduce considerable bias. Since we removed extreme outliers, RSD is more appropriate here, though we retain RMSE for comparison with other studies.

The two groups of values provide a convenient juxtaposition: $R^2$, with a scale of 0 to 1, estimates the variation explained by the independent variables. In contrast, RMSE/RSD represents the unexplained variation, with units in which the dependent variable is reported.

We also included MAE (the mean absolute error), a simple but useful measure. It is the average mean absolute difference between the predicted and actual values from the regression. Its derivative MAPE (the mean absolute percentage error) is the mean absolute difference expressed as a percentage of the actual value. Finally, we analysed the AIC (Akaike Information Criteria) to identify a model that most accurately fits the data without overfitting (i.e. it avoids identifying a regression that fits the existing data very well but at the expense of providing an optimum prediction of new data). It actively penalises any superfluous additional independent variables. The smaller the value, the better the model. Thus, an increase in the R-value(s) and a decrease in RMSE, RSE/RSD, MAE, MAPE, or AIC indicates an improvement in the fit of a regression model.

c. Assessment of the regression models in estimating metabolisable energy density

The variation in the number of laboratories contributing to the data in each of the four analytical regression models restricted our ability to determine the models' relative accuracy. However, our data included 359 records that contained the full range of analytes. We calculated the predicted ME density for each of the records of the four regression analytical models. The goodness of fit measurements, particularly the mean absolute error (MAE), gave unbiased estimates of the accuracy and precision of the four models. We used two further statistical tests to determine the differences in absolute residual values. The two tests were an ANOVA, as described above, and a test of equivalence using a multiple-sample TOST test provided in the InVivoStat software package (https://invivostat.co.uk/). Equivalence testing increases the confidence that two or more treatments (their means and 95% confidence intervals) are equivalent within stated bounds, in our case, about 2% rather than simply not significantly different as in an ANOVA [55].

d. Validation: the accuracy and precision of the regression analyses for future sets of data

Although it is common to describe regression equations as predictive, they only explain the data on which they are based and may be of doubtful accuracy for a predictive model. Investigators often overlook the subtle difference between the explanatory and predictive functions [56]. The change in function does not affect the regression coefficients themselves. The main effect is on goodness of fit and the magnitude of explained and unexplained variation: this can be a severe problem with small datasets. We tested the predictive quality of our regression in two ways.

The most comprehensive method was the k-fold cross-validation [57; 58] with a k value of 10: the analysis is provided in the R caret package. The procedure assumes that one can mimic a future set of data that complies with the characteristics of the current population. The existing data are divided into k portions or folds. The estimated ME density of each of the ten portions ('the future dataset') is obtained sequentially from predictive equations from the remaining nine data portions. This procedure produces ten estimates of one or more measurements of the goodness of fit. As a default, caret provides $R^2$, RMSE, and MAE as GOF indices. We extended the code to add adjusted $R^2$ and RSD (the code is shown in the Supplementary Material). We repeated the procedure ten times to improve the reliability of the values. Our mean values and their 95% confidence intervals were thus obtained from 100 estimates. We also obtained the predicted $R^2$ using the olsrr R package. This value is a simpler validation measure based on dividing the dataset into a training and evaluation group. After that, the procedure is similar to the k-fold cross-validation, but it only provides a single estimate of $R^2$.

## Results

### *Phase 1. Evidence that energy digestibility in pigs and rats is similar.*

The preliminary data assessment shown in the Supplementary Material showed good agreement in the energy digestibility of pigs and rats above 0.65D. Nineteen pairs of data below 0.65D were considered outliers and excluded. Figure 1 displays a Bland-Altman plot of the remaining 185 pairs of combined data sets above 0.65D. Although the bias differed significantly from zero with 95% confidence intervals of 0.011 to 0.020, the biological difference was negligible, with digestibility in pigs about 0.015 units greater than in rats. This bias amounts to about 2% of our data's typical digestibility of 0.83. The outer levels of agreement were -0.042 to 0.073. Our interim conclusion was that we could proceed and use pigs as a model for murine models if digestibility was greater or equal to 0.65.

### *Phase 2. The development of predictive estimates of dietary metabolisable energy (ME) density based on dietary chemical analysis.*

*Removal of unsuitable and anomalous data*

We excluded eight records before statistical analysis. Five records had a DE/GE ratio of less than 0.65 and were outside the range of agreement between pig and murine digestibility; one of the five had an unusual ME/DE value of less than 0.90. We deleted three records from a single laboratory from the NDF and NDFS groups because of doubts about the accuracy of the analytical data (the sum of the measured analytes substantially exceeded the dry matter).

*Removal of outliers*

The number of statistical outliers identified and excluded in the four datasets using the IF method was less than 2.5%: CF, 14 of 697 (2%); NDF, 18 of 845 (2.1%); NDFS, 12 of 498 (2.4%); and NDFSS, 5 of 372 records (1.3%) (Figure 2).

*Dataset descriptive statistics*

Table 1 shows the descriptive statistics for the four groups after outlier removal. The nutritional values encompass those of typical murine breeding, general-purpose, and maintenance diets. Removing outliers had a small effect on the analyte concentrations, with a mean difference of -0.40% (SD 0.78) of the values in the equivalent original dataset. We were concerned to see negative minimum values for Res1 and Res2 in a few records, some of which occurred in laboratories we considered to be highly competent. Negative residue values indicate an error in one or more measured analytical variables. However, we retained these records to avoid unintended bias since these analytical discrepancies might be present in other records but less visible.

*Correlation matrices*

We have provided the full matrices in the Supplementary Material. The correlation matrices are an objective method for deciding which pair of predictors to use in the 'drop-one' approach in regressing compositional data. While the correlations between ash and residues for the CF (-0.65) and NDF (-0.58) models support dropping ash or the residue, that is less so for the NDFS (0.17) and NDFSS models (0.41). In the latter two models, substituting ash and starch is statistically the best option but unhelpful when starch is an important predictor to study. We consider context to be an important consideration in our choice. Thus, for compatibility with previous studies and consistency across our models, we removed either ash or the calculated residue (NFE, NFC, Res1, or Res2) from all four dataset regressions. We refer to the options as ash-based and residue-based models. We describe our approach to choosing between the two models in the regression collinearity section below.

The matrices also provided other nutritionally interesting associations. There was a high correlation (0.98) between DE and ME density measurements across the four analytical datasets. Some authors included ME values calculated from DE, which inevitably influences the correlation. Nevertheless, there is undoubtedly a close relationship. The correlations between GE and DE or ME density were poor (0.29 to 0.47). The ether extract was consistently highly correlated with GE (0.79 to 0.92) but not DE or ME density. Both DE and ME were moderately negatively correlated with ash (-0.40 to -0.62), probably because ash is

the inverse of energy-containing organic matter. Energy digestibility (DE/GE) consistently showed a high negative correlation with crude or neutral detergent fibre (-0.67 to -0.77) and a moderately positive one with NFC or starch (0.59 to 0.64). The typically low levels of sugars had no effect.

Of the nutrient-nutrient correlations, only two exceed the cutoff of 0.8 to 0.9 associated with a possible collinearity problem [59]. These were crude protein and NFE (-0.80) in the crude fibre dataset and starch and ash (-0.81) in the NDFSS dataset. The remaining correlations ranged from -0.73 to 0.54, with a mean absolute value of 0.31 (SD 0.21).

*Regression analyses*

Unless otherwise stated, the comments refer to the regression models after removing outliers.

*Regression collinearity and variance inflation factors (VIF)*

In addition to the correlation matrices, we assessed the risk of collinearity of the regression coefficients using variance inflation factors (VIF). VIF values, and thus collinearity, were much lower in the ash-based than residue-based models (Tables 2a to 2d). The effect of VIF on the precision of a calculated coefficient was estimated from the CVse as described in the Materials and Methods. Despite the higher VIF values in the residue models, there was no statistical difference between the precision of the coefficients in the ash- or residue-based regressions across the four models (t-test with unequal variances, t=2.06, P=0.26, df=25). The complete data are shown in the Supplementary Material.

Although using the ash models avoids the issue of collinearity, it can produce coefficients that, while mathematically correct, appear biologically odd. For example, in the NDFS GE density model, starch has a coefficient of zero with P=0.54. Although the VIF in the residue-based models is high, the effect is moderated by the high $R^2$ values and the large sample number, as described in the Discussion. The adjusted $R^2$ for ME density regressions for the four models varies from 0.751 to 0.869, with 367 to 827 records. Thus, in this study, VIF values exaggerate the risk of detrimental collinearity. However, since using the residue model is not universally accepted, we have included both ash- and residue-based intercept regressions below. The predicted energy densities and GOF values are the same in both cases.

*Regression models*

In summary, we created intercept and non-intercept regressions for dietary GE and ME density from four datasets (CF, NDF, NDFS, and NDFSS). Tables 2a-d show the nutrients included in each. Each intercept regression included the ash- and residue-based alternatives.

*The effect of potential laboratory bias*

After removing outliers, the mean difference in Estimated Marginal Means (EMMs) between the laboratory groups and Lab Group 0 across the four datasets was 0.02 kJ/g (SD 0.28 kJ/g), which represents approximately 0.1% of the average dietary metabolisable energy (ME) density (Supplementary Material). We concluded we could ignore any Laboratory Group bias and analyse the data as a whole.

*Regression parameters*

*Intercept models*

Tables 2a-d show regression analyses on the four datasets from which outliers have been excluded. Removing outliers had a negligible effect on the coefficients (typically 1 to 3% at most). The coefficients for the individual intercept models shown in the tables require little comment. The intercept for most GE density models based on residues was not significantly different from zero. The exception was that for the NDFS regression: nevertheless, it was small (c. 2.7 kJ/g, CI 1.3 to 4.1). The intercepts for the ash models were substantial (generally 17 to 18 kJ/g) and significant ($P<0.001$).

The predictor coefficients for GE and ME density were highly significant (mainly $P<0.001$) with two exceptions: (a) starch in the GE and ME density NDFS Ash models ($P=0.54$, $P=0.16$), and (b) NDF in the GE density NDFSS Ash model ($P=0.053$). Although many ME density regressions have negative coefficients for crude fibre, we observed it only in the ash-based model (and the no-intercept model below). However, even when not negative, the ME coefficients for crude fibre were much lower than the biological estimate of 7.54 kJ/g (Table 3).

The age-related Phase AD coefficients for the GE density models were also consistently non-significant, close to zero, and unaffected by the model. In contrast, those for ME density models were highly significant and much larger. The average age-related coefficient shows that older animals (as defined in the Materials and Methods) absorb an additional 0.37 kJ/g from a diet than the younger grower/finisher animals. The amount increased from about 0.25 to 0.48 kJ/g across the four models.

The coefficients for ash and alternative residues (and their t values, though not shown here) appear as mirror images across comparable regression models, a phenomenon not commonly reported in the literature. This pattern occurs when any one of a pair of predictors is substituted and is unrelated to their coefficient in the correlation matrix. Its cause is beyond the scope of this paper.

The individual nutrient GE density values determined with the residue models were about 95% (SD 8.1%) of their respective theoretical values (Table 3). We ascribe the slight discrepancy to the contribution of energy density from the intercept. In the case of the residue-based ME density regressions, the very high intercept values made any attempt to relate the coefficients with theoretical values pointless. There was no similarity between the coefficients and theoretical nutrient GE or ME density in the ash-based regressions. The most reliable comparisons of regression and theoretical values are with the no-intercept models reported below.

*The goodness of fit (GOF) of the intercept models*

The GOF values for the intercept models are the same for the residue-based and ash-based models and are shown in Tables 4a (GE density regressions) and 4b (ME density regressions). Removing outliers had little effect on the GOF values for the GE density regressions. Goodness of fit progressively improved with model complexity. The predictors explain about 77 to 94% of the variation in GE density, and unexplained variation is low. For example, RSD averaged 0.23 kJ/g, progressively decreasing with more independent variables in the models. The unexplained variation in the models represents only 1-2% of the typical GE density of murine and pig diets (about 18.4 kJ/g).

Although we removed only about 2% of the records as outliers, there was a modest improvement in the GOF of ME density regressions. The various $R^2$ estimates improved by about 2% across the four groups, while RMSE and RSD values improved by 7% and MAE and MAPE by about 5%. Not surprisingly, considering the additional complexity of measurement and variation in the metabolisability of the fibre components, GOF values were poorer in the ME density prediction models. Nevertheless, the independent variables still account for about 75 to 87% of the variation in the dependent variable, which is good for most science-based studies [60]. We comment in the Discussion section on the source of the remaining variation. The RSD values for the ME density regressions were 50-80% higher than for GE regressions in the crude fibre and NDF models and over two-fold higher in the NDFS and NDFSS models. However, the MAE levels were low even in the more challenging ME prediction models and were only about 2-3% of the ME density of typical pig or murine diets (14.8 kJ/g). The MAE was highest with the crude fibre model, about 0.42 kJ/g (CI 0.40, 0.45), and this improved in the NDFSS model to about 0.29 kJ/g (CI 0.27, 0.31).

The GOF measurements across the four analytical models show that the crude fibre-based model is the weakest fit while the NDFSS model is the strongest. This trend is not just a

function of an increasing number of independent variables since AIC, which penalises such increases, substantially improves across the models. The NDFS model generated the largest single-step improvement in fit. Unfortunately, one must treat this comparison with caution since the number of laboratory groups (and hence potential variability) differs in the four models: we address this later when we analyse the GOF indices in 359 records that include all the analytes.

*No-intercept models*

The no-intercept models represent the net contribution of energy-containing nutrients to the regression. The merit of the no-intercept approach is that it generates biologically relevant coefficients, though they are still empirically based. However, the no-intercept model does not allow a meaningful coefficient for a factor like Phase: the outcome is a value tantamount to an intercept. Because of the expected insignificant effect of Phase on GE density in the intercept models, we have reported GE density coefficients in the no-intercept models on the whole data (Table 5a). In contrast, there was a significant Phase effect in the intercept-based ME density regressions, so we performed no-intercept ME density regressions for grower/finisher and adult animals separately (Table 5b).

The percentage differences in the comparable coefficients after removing outliers were negligible (typically 1%). The main exception was the crude fibre coefficient in the ME density regression: this was probably a consequence of the small coefficient values and the relatively large effect of a small difference after outlier removal.

The GE density regression coefficients were similar to those of Noblet and Van Milgen [(52)], and our theoretical estimates for crude protein, ether extract, and carbohydrates gathered from various publications (Table 3). The overall agreement with theoretical estimates was 101.2% (SD 3.4%). We could not find direct estimates of the ME density of individual nutrients, presumably because suitable experiments would be complex. However, we have calculated estimates of nutrient ME density from published data (Table 3). There were three points of interest: first, the regression values for the ‘available’ carbohydrates (NFE, NFC, Starch, and Sugars) and ether extract were extremely close to the calculated values (mean 0.98, SD 0.03). Second, the regression values were much lower than theoretical for the fibre constituents (CF, NDF). Finally, the crude protein value was c.20% higher from the regressions than theoretic values. Overall, the values were consistent with crude fibre primarily affecting protein absorption with energy losses integrated into the fibre coefficients.

*The goodness of fit (GOF) of the no-intercept models*

Outlier removal had little effect on the GOF values for GE density predictions (median improvement < 1%) but improved those for ME density (median c. 5-6%). Table 6 shows the GOF indices of the no-intercept models, which follow the same trend as the intercept models. The GOF values of the GE and ME density regressions generally improve with increasing model complexity. The difference between the CF and NDF ME density prediction models for grower/finisher pigs is an exception. However, the AIC values show that the overall improvement is not just an effect of increasing the number of predictors. The GOF of no-intercept models for ME density is poorer than the comparable intercept models. The effect is most evident in the estimates of mean absolute error, especially for the NDF and NDFS regressions. Despite the attractiveness of biologically relevant regression coefficients, the mean absolute error values indicate that the prediction of ME density is slightly less accurate than the intercept model.

*Assessment of the regression models in estimating metabolisable energy density*

Figure 3 shows the regression plots of measured ME density on predicted ME density for the four models. We regressed the data in this order since it is important to understand how well the calculated energy density value represents the measured value. It is also the orientation recommended by Piñeiro *et al.* [61]. In all four cases, the slopes were 1.00 and the intercept zero: these values demonstrate a perfect agreement between mean predicted and actual values across the range of values studied. The 95% confidence interval of the prediction of the mean response (confidence interval, CI) in these regressions is small and not clearly evident in the figures. Although the CI is tapered, we have calculated its mean value as ± 0.5 kJ/g across all four regressions. Perhaps the best overall practical measure of the predictive accuracy of our regressions is the 95% confidence interval of the prediction of a single value (prediction interval, PI). It provides the range within which the true value of a prediction is likely to fall 95% of the time. The prediction intervals are almost linear over the range of ME density values, so we can reasonably express them as a single value for each of the four analytical models. The prediction intervals were ± 1.03 kJ/g (CF), ± 0.97 kJ/g (NDF), ± 0.79 kJ/g (NDFS), and ± 0.71 kJ/g (NDFSS). These values represent a deviation from the average predicted ME value for the four models (14.8 kJ/g) of ± 7.0, 6.6, 5.3, and 4.8%, respectively. As expected, the absolute PIs are about twice the RMSE/RSD values.

*Choosing an optimum cost-effective regression model to estimate the metabolisable energy density.*

We obtained an unbiased estimate of the relative accuracy and precision of the four analytical regression models by analysing the 359 records present in all four datasets. Table 7 shows the GOF indices used. Notably, these regressions and GOF measurements come from a narrow range of laboratories with less diversity of factors such as animal breed, technique, and measurement accuracy than records in the whole dataset. Consequently, it is unsurprising that the coefficients and GOF measures within these datasets are generally more precise than the equivalent whole dataset. The major exception is the regression analysis for the NDFSS dataset, in which the coefficients and GOF measurements in the two were very similar. However, the 359-records dataset included a substantial proportion of the NDFSS dataset. As with the primary data (Table 4), the GOF estimates improved with the complexity of the model.

Since the mean actual and predicted ME densities of the four regression models using the common dataset are the same (14.43 kJ/g) and the average of the residuals zero, we compared the accuracy and precision of the four models by analysing the absolute residuals. Figure 4 shows the distribution of the absolute errors as a raincloud plot. A preliminary examination for normality with a Q-Q plot (not shown) and the Shapiro-Wilk test showed that the regressions' absolute residuals were not normally distributed ($W=0.93$, $P<0.001$). Consequently, we used the nonparametric Kruskal-Wallis test, which showed that the overall difference between the regressions' absolute residuals was not statistically significant at the 5% level ($P=0.11$). No further *ad hoc* testing was deemed necessary. The InVivoStat Equivalence TOST Test showed that the four sets of absolute residuals are equivalent at the 5% level within ± 0.1 kJ/g. Thus, apart from the few outlying absolute residuals with the CF model, there seems to be little difference in the accuracy of the four regression models in predicting the ME density of individual records.

Finally, we obtained estimates of analysis costs for the analytes in the four regression models. Since these will change over time, we have expressed them relative to that of the CF model. The relative costs for the four analytical packages in 2023 were: CF, 1; NDF, 1.4; NDFS, 1.9; and NDFSS, 2.2. Despite the higher costs, the improvement in MAE over the four models for the global and common data is less than 0.15 kJ/g, which is only about 1% of the typical ME density.

*The effectiveness of the four regression models in estimating metabolisable energy density in future datasets*

Table 8 shows the 10-fold cross-validation data for the prediction of ME density in the four analytical models. The adjusted $R^2$, RSD, RMSE, and MAE values indicate negligible overfitting since they are little different from the equivalent values in Table 4. This observation suggests that the regressions provided in this paper can accurately predict the ME density of additional future datasets, providing the new data's nutrient levels and energy densities fall within the ranges shown in the descriptives. This assertion holds even if one takes the most unfavourable 95% confidence interval values. For example, in some future comparable datasets, the predictors in the crude fibre model would typically account for about 75% of the variation in the ME density prediction, and the MAE would be no more than ± 0.43 kJ/g. The improvement in prediction and GOF values from the crude fibre to NDFSS datasets follows a similar pattern to the existing regressions. Thus, the three NDF-based regressions would improve predictive quality over crude fibre, although the biological effect is small. The simpler predictive $R^2$ values were similar to those of the adjusted $R^2$.

*Comparison of regression equations with previously published ones.*

Although investigators have developed numerous regression equations to estimate ME density in pigs, many are based on only a few nutrients. We collated several regression equations developed to predict the ME density of pig diets using a more comprehensive range of nutrients similar to our selection (Table 9). The comparatively large intercept values in the ME-density intercept models make applying meaningful comparisons for the nutrient coefficients difficult. That issue is not a problem for those from the no-intercept models. While the crude fibre coefficient is often negative, that is not axiomatic, though its value is close to zero. Each regression provides a similar predicted ME density when applied to the analytical data from the grower/finisher subset of the 359 'common records' dataset. As one might expect, the GOF values on regressions based on a single laboratory are generally slightly better than those from this study. Morgan's RSD/RMSE values are much higher, possibly because the regressions are based on the combined data from ingredients and complete diets. We return to this information in the Discussion.

## Discussion

We have developed a series of predictive equations for the metabolisable energy (ME) density of murine diets from their nutrient analysis. There were insufficient suitable data from murine

studies, and we relied on data from pig studies. As Figure 1 shows, there is good agreement in digestibility between pigs and murine models above 0.65D, and several studies confirm that the similarity extends to metabolisability. Basing equations on pig data has major advantages. There is a large amount of published data from a wide range of laboratories, and the dietary nutrient levels and ingredients are those often seen in murine diets.

Variable selection was primarily dictated by analytical packages routinely reported by manufacturers for diets for laboratory rats and mice [(1)]. The most widely used is the proximate analysis, though NDF sometimes replaces crude fibre since it is considered a better estimate of fibre. Although routine pig or murine diet analysis rarely includes starch and sugars, they were part of the analytical packages in many publications. We included them in our datasets and regression models to test if their inclusion in routine nutrient analysis might be beneficial in providing more accurate and reliable estimates of energy density and sufficiently so to warrant the additional cost.

With compositional data such as ours, where the values of potential predictors should total 100, dropping one predictor is often used to avoid multicollinearity [(59)]. We decided to choose between ash and a residue, such as NFE or NFC, based on the output from the correlation matrices, our desire to be consistent across the models, and compatibility with other studies. There is no consistency in the literature (see, for example, Table 9). One can choose between ash or residue-based models from a statistical and biological perspective. While residue-based regressions tend to give more "physiologically meaningful" coefficients [(22)], they increase collinearity and result in higher VIF values for most coefficients. Although a high VIF may be associated with increased coefficient variance and imprecision, there is decreased risk with large sample numbers and high $R^2$ values [(59; 65; 66; 67; 68)]. Our estimation of CVse in the ash- and residue-based regressions and p-values confirmed no detrimental effect of collinearity on the precision of the coefficients. From a biological perspective, although ash could affect energy density by increasing the endogenous losses of protein and fat [(69; 70)], the evidence suggests that any effects are small or non-existent in pigs [(71; 72)] but perhaps significant in poultry [(73)]. We believe both ash- and residue-based regressions are satisfactory, and we include both since each has merits that are ultimately best assessed by the reader.

An equation generated from one data source may not apply more generally because of the many differences in the animals, environment, and methodology used [(24)]. We sought to overcome this limitation by selecting data from numerous studies encompassing various pig breeds, diet preparations and types, nutrient levels, environmental conditions, and technical

practices. By analysing data from such a broad spectrum, we mitigate the risks of overfitting and shrinkage [56; 74] and improve reproducibility [75], ensuring our models are more generalisable and reflective of real-world variability [76]. This approach significantly enhances the reliability of our predictive models when applied to future datasets. Choi [30], Sung [29], and their colleagues have taken a similar approach but with fewer samples and predictors and without equivalent prediction equations for dietary ME density.

This diverse background comes with a penalty, though we consider that the utilitarian nature of our regressions outweighs the disadvantages. The primary 'disadvantage' is that these other variable background factors increase the unattributed variation, resulting, for example, in slightly poorer adjusted $R^2$, RSD, and MAE values than one might hope. Nevertheless, the adjusted R2 values are still good, ranging from 0.75 to 0.87 for the four models [60]. The typical RSD from carefully conducted, single-site studies by Noblet is about 0.34 kJ ME/g for a crude fibre regression and about 0.30 kJ ME/g for the NDF one (Table 9). In our 'global' equations, the values were higher at 0.52 and 0.46 kJ/g for the crude fibre and NDF models. When we restricted the data to 359 sets common to all models and from fewer sources, the RSD values were about midway between the single site and 'global' models, at 0.45 and 0.37 kJ/g, respectively. Two related studies by Bulang and Rodehutscord [63] and Grümpel-Schlüter and colleagues [64] deviate from our expectations and produce unusually low RMSE values (median c. 0.25 kJ/g) despite a diverse data source. The divergence may result from calculating each diet's ME density from an equation recommended by the GfE [19; 20] rather than measurement. This approach inevitably decreases variation.

Investigators have used intercept and non-intercept regressions to predict energy density from dietary chemical constituents (Table 9). Removal of the intercept slightly decreases accuracy, as can be seen when comparing the values predicted with Noblet's (E) and our data (F) in Table 9 by the two regression types. The GOF was similar in both regression types.

The no-intercept models have some advantages: they avoid the issue of ash- or residue-based regressions discussed above, and unlike those in the intercept models, one can apply their coefficients to diets and ingredients alike [43]. Moreover, excluding intercepts, which are large contributors to the dependent variable, may provide "physiologically relevant" coefficients [22] that may appeal to some investigators. We use both forms of regression, with the intercept model providing the definitive prediction and the no-intercept model providing a (net) measure of the contribution of the energy-yielding nutrients to the total energy density [43].

The intercepts and nutrient coefficients in Table 2 apply to both grower/finisher and adult

animals. Adding a growth phase to the regressions is a powerful feature and accommodates data consolidation from the grower/finisher and adult animals in the regression. It defines the additional energy that should be added to the dependent variable for adult animals to reflect their greater energy absorption and metabolisability [62; 77; 78]. However, the ages of animals in most murine research and regulatory studies are probably best represented by grower/finisher pigs and the base prediction equations, even though the age effect may exist in older murine models [79]. Determining the effect of the growth phase on predicted ME density is only possible with the intercept models.

Though it is unlikely one would predict GE density from nutrient analyses, we have included GE density regressions here as a test of the regression process's effectiveness when applied to ME density predictions. Since the GE coefficients in the four models are affected by the intercept size, the best comparison is with the no-intercept regressions in Table 5a. Across the models, there is good agreement with the published values for simple analytes such as crude protein, ether extract, starch, and sucrose (Table 3) from the residue-based regressions, though not those based on ash. Assessing the accuracy of GE density coefficients for crude or neutral detergent fibre is more complex. Their energy content depends on their fibre composition, which varies with the ingredient [80; 81; 82; 83; 84; 85; 86]. Cellulose and hemicellulose, two of the main components in fibre, have values of about 16-18 kJ/g, while lignin ranges from 21-30 kJ/g [87; 88; 89; 90; 91]. Fibre may also include small amounts of protein (c.23-24 kJ/g).

Although obtaining reliable estimates of the GE density of nutrients from regressions is reassuring despite the empirical nature of the regressions, the study's primary purpose was to predict ME density. In contrast to the nutrient coefficients for GE density, there is considerable variation in comparable coefficients in this study's four ME density models and other published equations (Table 9). The exceptionally high intercept values contribute to much of the variation, especially in those regressions that include residues (NFE, NFC, Res1 or Res2). Overall, there is no correspondence in the intercept regressions between theoretical ME density values and those obtained from the coefficients. Nevertheless, despite a large variation in intercept and coefficient values between models, the different regression equations generally give predicted ME density values close to the actual values.

The no-intercept regressions allow the comparison of regression-derived nutrient coefficients, unaffected by the large and variable intercepts, with our theoretical estimates for ME density (Table 3). The fibre regression coefficients are much lower than calculated ME densities (CF, c.300% lower, and NDF 65% lower). In contrast, those for protein and ether extract are about

10-20% higher. These results are consistent since dietary fibre is known to increase energy loss from protein and fat [92; 93; 94], and the nature of the regression analysis allocates these losses to fibre rather than the protein and ether extract coefficients. There is good agreement for the soluble carbohydrates (0 to 5% difference) consistent with the absence of a significant effect of fibre on starch absorption [92; 95; 96].

The GOF values showed that our regression models were consistently good and improved with increasing characterisation of the diet's nutrient composition. The GOF measurements were inevitably poorer with the ME density regressions than those for GE density, reflecting increased experimental variation and errors associated with biological studies. Internal validation based on predicted $R^2$ and k-fold GOF measurements confirms that our regression equations explain existing relationships between energy density and the chemical analytes well and are suitable for future predictions without exhibiting overfitting. We are confident that the first three models provide generalised predictions suitable for widespread practical use. However, data from the Noblet group [24; 39; 62; 97] dominated the NDFSS model, contributing about 87% of the records. This may lead to some overfitting, though less than using data from a single study.

Although GOF indices such as RMSE, RSD or MAE reflect errors in the regression model, they apply to populations rather than individual samples. While this expression of error may be suitable for a diet manufacturer providing an illustrative ME density as a guideline, most investigators should be concerned with the prediction error in a single diet sample rather than that averaged over a population of samples. Here, the PI is the best estimate of uncertainty when new predicted values are required. Unfortunately, investigators often ignore the effect of the PI on a study. As with the MAE, PI improves across the four models but only by about 0.3 kJ/g. One should assume the PI could be up to ± 1 kJ/g of the predicted value.

External validation of prediction equations is important, although often overlooked. Its significance is particularly well-recognised in clinical studies [98]. We achieved some benefits of external validation by comparing the actual ME density of a test diet with values predicted from our equations and other published equations (Table 9). While this step does not replace traditional external validation, it still reinforces the validity of the equations. There was good agreement, though it was slightly better for the intercept than the no-intercept model.

In selecting the best regression model for routine use, one should consider the analysis's complexity, cost, and GOF. We found no evidence from either the Kruskal-Wallis or equivalence testing that any of the four analytical models was better at predicting ME density.

However, there was a slight improvement in GOF values with greater complexity. The CF model seems adequate for murine models, especially since murine laboratory diet analysis commonly includes the necessary predictors. Although NDF-based models improve the prediction error, they cost up to twice that of the CF model analysis. For some studies, these increases in cost for such little gain might be cost-effective, though unlikely.

There are three constraints to using our equations. First, although diet manufacturers provide typical nutrient estimates, we recommend the investigator obtain chemical analysis data on the diet used for optimal accuracy. Second, the regressions may not apply to purified diets for which the Atwater Factors remain relevant [11]. Third, their accuracy may decline with very high or low levels of individual nutrients, though such values would be uncommon for murine diets (or complete diets for pigs).

In developing our regressions, we have accumulated more data records than in previous publications and subjected them to rigorous statistical analysis. Possibly, for the first time, we have compared the relative merits of predictions based on several analytical packages in a single study and evaluated their cost-effectiveness. We have also provided realistic estimates of the possible prediction error when applied to diet samples or an individual batch. Our inability to validate these predictive equations with murine data remains a weakness and indicates the need for further reliable ME density measurements in rats and mice. We know of only one study that examined the application of a pig equation to natural-ingredient murine diets to predict ME density. Bielohuby and colleagues [11] found the equation (unattributed, but Kirchgessner and Roth's $ME_{BFS}$ equation [25] shown in Table 9) was more accurate than Atwater Factors for natural-ingredient rat diets. However, the eleven diets tested are too few to validate their use confidently.

Although near-infrared spectroscopy (NIRS) has been suggested as an alternative to predict dietary energy density, it is not straightforward. It is easily affected by small environmental and sample changes, may require several hundred calibration samples, and the equipment is expensive [99]. Nor do we consider its accuracy for energy density predictions better than ours (cf Noel and colleagues [100]).

Although several previous studies (Table 9) predict the ME density of pig diets from chemical analysis, this study provides more robust validated prediction equations, with standard combinations of the most common feed analytes as predictors. Several of the studies use only calculated ME density and no-intercept models. The regressions are from single laboratories, include fewer records, and inevitably underestimate prediction errors when used widely. We

used more extensive measures of goodness of fit and validation to ensure that the regressions would maintain their robustness when applied to new data. All four regression models give good GE and ME density predictions that can be applied to murine diets. While the quality of fit to the data improves slightly with increased numbers or refinement of predictors, analysis costs increase substantially. Moreover, the differences in prediction errors are small relative to the predicted values. Thus, the crude fibre equation seems an adequate practical approach to determining ME density. If one requires higher precision and accuracy than shown here, one must measure ME density *in vivo*.

**Acknowledgements**

We are grateful for further analytical data from Jean Noblet and his comments on our approach to the study. Many other investigators also provided helpful comments and expanded on their published data, particularly Su Lee and Brian Kerr. Finally, GT wishes to thank the staff of the Oxford University libraries, particularly for their efforts to maintain service during the COVID-19 pandemic.

**Financial support**

This research received no specific grant from any funding agency, commercial or not-for-profit sectors. AS is employed by ssniff Spezialdiaeten but has received no direct payments to support this study.

**Conflict of interest**

None

**Ethical statement**

There are no ethical issues in this study.

**Author contributions**

GT – Study design, data collection, statistical analysis, manuscript preparation.

AS – Study design, manuscript preparation and review.

TG, LS – Guidance on statistical analysis and using the R programming language, manuscript review.

## References

1. Tobin G, Schuhmacher A (2021) Laboratory animal nutrition in routine husbandry, and experimental and regulatory studies. In *Handbook of Laboratory Animal Science,* 4th ed., pp. 269-311 [J Hau and S Schapiro, editors]. Boca Raton, FL: CRC Press.

2. Tschop MH, Speakman JR, Arch JR *et al.* (2012) A guide to analysis of mouse energy metabolism. *Nat Methods* **9**, 57-63.

3. Rozman J, Klingenspor M, Hrabě de Angelis M (2014) A review of standardized metabolic phenotyping of animal models. *Mamm Genome* **25**, 497-507.

4. McLean JA, Tobin G (1987) *Animal and human calorimetry.* Cambridge, UK: Cambridge University Press.

5. Elvert R, Wille A, Wandschneider J *et al.* (2013) Energy loss via urine and faeces – a combustive analysis in diabetic rats and the impact of antidiabetic treatment on body weight. *Diabetes, Obesity and Metabolism* **15**, 324-334.

6. Armitage G, Hervey GR, Rolls BJ *et al.* (1983) The effects of supplementation of the diet with highly palatable foods upon energy balance in the rat. *J Physiol* **342**, 229-251.

7. Atwater W (1900) *Discussion of the terms digestibility, availability, and fuel value*. *12th Annual Report of Storrs Agr Exp Sta*. Storrs, CT.

8. Merrill AL, Watt BK (1955) *Energy value of foods: basis and derivation.* Washington, DC: US Department of Agriculture.

9. Sanchez-Pena MJ, Marquez-Sandoval F, Ramirez-Anguiano AC *et al.* (2017) Calculating the metabolizable energy of macronutrients: a critical review of Atwater's results. *Nutr Rev* **75**, 37-48.

10. Maynard LA (1944) The Atwater system of calculating the caloric value of diets. *J Nutr* **28**, 443-452.

11. Bielohuby M, Bodendorf K, Brandstetter H *et al.* (2010) Predicting metabolisable energy in commercial rat diets: physiological fuel values may be misleading. *Br J Nutr* **103**, 1525-1533.

12. Schneider BH (1947) *Feeds of the World : their digestibility and composition.* Blacksburg, WV: Agricultural Experiment Station, West Virginia Univ.

13. Crampton EW, Harris LE (1969) *Applied animal nutrition : the use of feedstuffs in the formulation of livestock rations*. 2nd ed. San Francisco, CA: Freeman and Company.

14. Flachowsky G, Kirchgessner M (1998) The energetic feed evaluation in Germany. *Archiv für Tierernährung* **51**, 111-125.

15. Eeckhout W, Moermans R (1981) Prediction of the energy value of commercial pig feeds, according to the various energy systems used in the European Community that are based on simple analytical parameters of the feed. *Anim Feed Sci Technol* **6**, 367-375.

16. Schiemann R, Nehring K, Hoffmann L *et al.* (1971) *Energetische Futterbewertung und Energienormen.* Berlin, GDR: VEB Deutscher Landwirtschaftsverlag.

17. Jentsch W, Beyer M, Chudy A (2001) Die Rostocker Arbeiten zur energetischen Futterbewertung und zum Energiebedarf landwirtschaftlicher Nutztiere. 2. Mitteilung: Zur Weiterentwicklung des Rostocker Futterbewertungssystems. *Übers Tierernährg* **29**, 1-44.

18. Jentsch W, Chudy A, Beyer M (2000) Die Rostocker Arbeiten zur energetischen Futterbewertung und zum Energiebedarf landwirtschaftlicher Nutztiere. 1. Mitteilung: Historischer Rückblick und die Arbeiten in Rostock zur energetischen Futterbewertung. *Übers Tierernährg* **28**, 133-182.

19. GfE (2006) *Empfehlungen zur Energie- und Nährstoffversorgung von Schweinen*. Frankfurt am Main, DE.

20. GfE (2008) Prediction of metabolisable energy of compound feeds for pigs. *Proc Soc Nutr Physiol* **17**, 199-204.

21. Morgan DJ, Cole DJA, Lewis D (1975) Energy values in pig nutrition: II. The prediction of energy values from dietary chemical analysis. *J Agric Sci* **84**, 19-27.

22. Just A, Jørgensen H, Fernández JA (1984) Prediction of metabolizable energy for pigs on the basis of crude nutrients in the feeds. *Livest prod sci* **11**, 105-128.

23. Morgan CA, Whittemore CT, Phillips P *et al.* (1987) The prediction of the energy value of compounded pig foods from chemical analysis. *Anim Feed Sci Technol* **17**, 81-107.

24. Noblet J, Perez JM (1993) Prediction of digestibility of nutrients and energy values of pig diets from chemical analysis. *J Anim Sci* **71**, 3389-3398.

25. Kirchgessner M, Roth FX (1983) Schätzgleichungen zur Ermittlung des energetischen Futterwertes von Mischfuttermitteln für Schweine. *Z Tierphysiol Tierernahr Futtermittelkd* **50**, 270-275.

26. Furuya S, Takahashi S, Kameoka K (1981) A comparative study of the digestibility by pigs, chickens and rats using identical rations. *Jpn J Zootech Sci* **52**, 459-466.

27. Nehring K, Hoffmann L, Schiemann R *et al.* (1963) Die energetische Verwertung der Futterstoffe. 6. Mitteilung. Die energetische Verwertung der Kraftfutterstoffe durch Rinder, Schafe, Kaninchen, Schweine und Ratten Tierartenvergleich. *Archiv für Tierernaehrung* **13**, 193-213.

28. Jørgensen H, Lindberg JE (2006) Prediction of energy and protein digestibility in pig feeds using growing rats as a model. *Anim Feed Sci Technol* **127**, 55-71.

29. Sung JY, Kim BG (2021) Prediction equations for digestible and metabolizable energy concentrations in feed ingredients and diets for pigs based on chemical composition. *Asian-Australas J Anim Sci* **34**, 306-311.

30. Choi H, Sung JY, Kim BG (2020) Neutral detergent fiber rather than other dietary fiber types as an independent variable increases the accuracy of prediction equation for digestible energy in feeds for growing pigs. *Asian-Australas J Anim Sci* **33**, 615-622.

31. R Core Team (2023) *A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing.

32. Altman DG, Bland JM (1983) Measurement in medicine: the analysis of method comparison studies. *J Roy Stat Soc Ser D (Statistician)* **32**, 307-317.

33. Giavarina D (2015) Understanding Bland Altman analysis. *Biochem medica* **25**, 141-151.

34. Ludbrook J (2010) Confidence in Altman-Bland plots: a critical review of the method of differences. *Clin Exp Pharmacol Physiol* **37**, 143-149.

35. Morgan CA, Whittemore CT (1982) Energy evaluation of feeds and compounded diets for pigs — A review. *Anim Feed Sci Technol* **7**, 387-400.

36. Kromann RP (1967) A mathematical determination of energy values of ration ingredients. *J Anim Sci* **26**, 1131-1134.

37. Young LG, Ashton GC, Smith GC (1977) Estimating the energy value of some feeds for pigs using regression equations. *J Anim Sci* **44**, 765-771.

38. Kim JW, Koo B, Nyachoti CM (2017) Digestible, metabolizable, and net energy of camelina cake fed to growing pigs and additivity of energy in mixed diets. *J Anim Sci* **95**, 4037-4044.

39. Noblet J, Fortune H, Dupire C *et al.* (1990) Valeur nutritionnelle de treize matières premières pour le porc en croissance. 1. Teneurs en énergie digestible métabolizable et nette. Conséquences du choix du système énergétique. *Journées de Recherche Porcine en France* **22**, 175-184.

40. Noblet J, Henry Y (1993) Energy evaluation systems for pig diets: a review. *Livest prod sci* **36**, 121-141.

41. Kerr BJ, Easter RA (1995) Effect of feeding reduced protein, amino acid-supplemented diets on nitrogen and energy balance in grower pigs. *J Anim Sci* **73**, 3000-3008.

42. Tobin G, Carpenter KJ (1978) The nutritional value of dry bean (*Phaseolus vulgaris*); a literature review *Nutr Abstr Rev Ser A Hum Exp* **48**, 919-936.

43. Noblet J, Wu S-B, Choct M (2022) Methodologies for energy evaluation of pig and poultry feeds: A review. *Anim Nutr* **8**, 185-203.

44. Wu S-B, Choct M, Pesti G (2020) Historical flaws in bioassays used to generate metabolizable energy values for poultry feed formulation: a critical review. *Poult Sci* **99**, 385-406.

45. Kong C, Adeola O (2014) Evaluation of amino acid and energy utilization in feedstuff for swine and poultry diets. *Asian-Australas J Anim Sci* **27**, 917-925.

46. Harrell FE (2001) *Regression modeling strategies with applications to linear models, logistic and ordinal regression, and survival analysis*, *Springer Series in Statistics*. New York, NY: Springer.

47. Riley RD, Ensor J, Snell KIE *et al.* (2020) Calculating the sample size required for developing a clinical prediction model. *BMJ* **368**, m441.

48. Liu FT, Ting K, Zhou Z-H (2009) Isolation Forest. In *2008 Eighth IEEE International Conference on Data Mining*. Pisa, Italy: IEEE.

49. Belsley DA, Kuh E, Welsch RE (1980) *Regression diagnostics: Identifying influential data and sources of collinearity*. New York: John Wiley & Sons.

50. Eisenhauer JG (2003) Regression through the Origin. *Teaching Statistics* **25**, 76-80.

51. Tay-Zar A-C, Wongphatcharachai M, Srichana P *et al.* (2024) Prediction of net energy of feeds for broiler chickens. *Anim Nutr* **16**, 241-250.

52. Noblet J, Van Milgen J (2013) Energy and energy metabolism in swine. In *Sustainable Swine Nutrition*, pp. 23-57 [LI Chiba, editor]. Ames, IA: Wiley-Blackwell.

53. Leach LF, Henson RK (2007) The use and impact of adjusted $R^2$ effects in published regression research. *MLRV* **33**, 1-11.

54. Kozak A, Kozak RA (1995) Notes on regression through the origin. *For Chron* **71**, 326-330.

55. Anderson-Cook CM, Borror CM (2016) The difference between “equivalent” and “not different”. *Quality Engineering* **28**, 249-262.

56. Yarkoni T, Westfall J (2017) Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspect Psychol Sci* **12**, 1100-1122.

57. Stone M (1974) Cross-validatory choice and assessment of statistical predictions. *J Roy Stat Soc Ser B (Stat Method)* **36**, 111-133.

58. Geisser S (1975) The predictive sample reuse method with applications. *JASA* **70**, 320-328.

59. Mason CH, Perreault Jr WD (1991) Collinearity, power, and interpretation of multiple regression analysis. *J Mark Res* **28**, 268-280.

60. Gupta A, Stead TS, Ganti L (2024) Determining a meaningful R-squared value in clinical medicine. *Academic Medicine & Surgery*.

61. Piñeiro G, Perelman S, Guerschman JP *et al.* (2008) How to evaluate models: Observed vs. predicted or predicted vs. observed? *Ecol Model* **216**, 316-322.

62. Le Goff G, Noblet J (2001) Comparative total tract digestibility of dietary energy and nutrients in growing pigs and adult sows. *J Anim Sci* **79**, 2418-2427.

63. Bulang M, Rodehutscord M (2009) Development of equations for predicting metabolisable energy concentrations in compound feeds for pigs. *Arch Anim Nutr* **63**, 442-454.

64. Grümpel-Schlüter A, Berk A, Schäffler M *et al.* (2021) Evaluation of an equation for predicting metabolisable energy concentration in compound feeds for pigs. *Arch Anim Nutr* **75**, 251-262.

65. Grewal R, Cote JA, Baumgartner H (2004) Multicollinearity and measurement error in structural equation models: Implications for theory testing. *Mark Sci* **23**, 519-529.

66. Feng W, Mullen MR, Sheng S (2017) Evaluating the impact of multicollinearity on regression. *Am J Bus Res* **9**.

67. O'Brien RM (2007) A caution regarding rules of thumb for variance inflation factors. *Qual Quant* **41**, 673-690.

68. Das S, Chatterjee S (2011) Multicollinearity problem–root cause, diagnostics and way outs. *Diagnostics and Way Outs (April 29, 2011)*.

69. Selle PH, Ravindran V, Cowieson AJ *et al.* (2010) *Phytate and phytase*: CABI Publishing: Wallingford, UK.

70. Stroebinger N, Rutherfurd SM, Henare SJ *et al.* (2021) Fatty acids from different fat sources and dietary calcium concentration differentially affect fecal soap formation in growing pigs. *J Nutr* **151**, 1102-1110.

71. Jørgensen H, Jakobsen K, Eggum BO (1992) The influence of different protein, fat and mineral levels on the digestibility of fat and fatty acids measured at the terminal ileum and in faeces of growing pigs. *Acta Agriculturae Scandinavica, Section A - Animal Science* **42**, 177-184.

72. Espinosa CD, Torres LJ, Velayudhan DE *et al.* (2022) Ileal and total tract digestibility of energy and nutrients in pig diets supplemented with a novel consensus bacterial 6-phytase variant. *J Anim Sci* **100**.

73. Cowieson A, Ruckebusch J-P, Sorbara J *et al.* (2017) A systematic view on the effect of microbial phytase on ileal amino acid digestibility in pigs. *Anim Feed Sci Technol* **231**, 138-149.

74. Babyak MA (2004) What you see may not be what you get: A brief, nontechnical introduction to overfitting in regression-type models. *Psychosom Med* **66**, 411-421.

75. Voelkl B, Vogt L, Sena ES *et al.* (2018) Reproducibility of preclinical animal research improves with heterogeneity of study samples. *PLoS Biol* **16**, e2003693.

76. Hastie T, Tibshirani R, Friedman J (2009) *The Elements of Statistical Learning. Data Mining, Inference, and Prediction*, *Springer Series in Statistics.* New York, NY: Springer

77. Lowell JE, Liu Y, Stein HH (2015) Comparative digestibility of energy and nutrients in diets fed to sows and growing pigs. *Arch Anim Nutr* **69**, 79-97.

78. Noblet J, Shi XS (1993) Comparative digestibility of energy and nutrients in growing pigs fed ad libitum and adults sows fed at maintenance. *Livest prod sci* **34**, 137-152.

79. Itoh H, Kaneko M, Ohshima S *et al.* (2002) Effect of low protein and low energy diet on physiological status and digestibility of F344 rats. *Exp Anim* **51**, 485-491.

80. Mongeau R, Brassard R (1982) Determination of neutral detergent fiber in breakfast cereals: Pentose, hemicellulose, cellulose and lignin content. *J Food Sci* **47**, 550-555.

81. Norman AG (1935) The composition of crude fibre. *J Agric Sci* **25**, 529-540.

82. Stallcup OT (1958) Composition of crude fiber in certain roughages. *J Dairy Sci* **41**, 963-968.

83. Hindrichsen IK, Kreuzer M, Madsen J *et al.* (2006) Fiber and lignin analysis in concentrate, forage, and feces: Detergent versus enzymatic-chemical method. *J Dairy Sci* **89**, 2168-2176.

84. Miron J, Yosef E, Ben-Ghedalia D (2001) Composition and in vitro digestibility of monosaccharide constituents of selected byproduct feeds. *J Agric Food Chem* **49**, 2322-2326.

85. Jaworski NW, Lærke HN, Bach Knudsen KE *et al.* (2015) Carbohydrate composition and in vitro digestibility of dry matter and nonstarch polysaccharides in corn, sorghum, and wheat and coproducts from these grains1. *J Anim Sci* **93**, 1103-1113.

86. Knudsen KEB (1997) Carbohydrate and lignin contents of plant materials used in animal feeding. *Anim Feed Sci Technol* **67**, 319-338.

87. Kaltschmitt M, Hartmann H, Hofbauer H (2009) *Energie aus Biomasse. Grundlagen, Techniken und Verfahren*. 2nd ed. Berlin, DE: Springer.

88. Demirbaş A, Demirbaş AH (2004) Estimating the calorific values of lignocellulosic fuels. *Energy explor exploit* **22**, 135-143.

89. Demirbaş A (2003) Relationships between heating value and lignin, fixed carbon, and volatile material contents of shells from biomass products. *Energy sources* **25**, 629-635.

90. Esteves B, Sen U, Pereira H (2023) Influence of chemical composition on heating value of biomass: A review and bibliometric analysis. *Energies* **16**, 4226.

91. Kienzle E, Schrag I, Butterwick R *et al.* (2001) Calculation of gross energy in pet foods: new data on heat combustion and fibre analysis in a selection of foods for dogs and cats. *J Anim Physiol Anim Nutr (Berl)* **85**, 148-157.

92. Agyekum AK, Nyachoti CM (2017) Nutritional and metabolic consequences of feeding high-fiber diets to swine: A review. *Engineering* **3**, 716-725.

93. Dégen L, Halas V, Babinszky L (2007) Effect of dietary fibre on protein and fat digestibility and its consequences on diet formulation for growing and fattening pigs: A review. *Acta Agric Scand A Anim Sci* **57**, 1-9.

94. Zhang S, de Vries S, Gerrits WJJ (2024) Quantifying the effects of dietary fibres on protein digestibility in pigs - A review. *Anim Feed Sci Technol* **308**, 115864.

95. Wilfart A, Montagne L, Simmins PH *et al.* (2007) Sites of nutrient digestion in growing pigs: Effect of dietary fiber. *J Anim Sci* **85**, 976-983.

96. Just A (1982) The influence of ground barley straw on the net energy value of diets for growth in pigs. *Livest prod sci* **9**, 717-729.

97. Shi X, Noblet J (1993) Digestible and metabolizable energy values of ten feed ingredients in growing pigs fed ad libitum and sows fed at maintenance level; comparative contribution of the hindgut. *Anim Feed Sci Technol* **42**, 223-236.

98. Steyerberg EW (2019) *Clinical prediction models: a practical approach to development, validation, and updating*. 2nd ed. Cham, CH: Springer.

99. Maduro Dias C, Nunes H, Borba A (2024) Near-infrared spectroscopy in animal nutrition: Historical insights, technical principles, and practical applications. *Analytica* **5**, 481-498.

100. Noel SJ, Jørgensen HJH, Knudsen KEB (2022) The use of near-infrared spectroscopy (NIRS) to determine the energy value of individual feedstuffs and mixed diets for pigs. *Anim Feed Sci Technol* **283**, 115156.

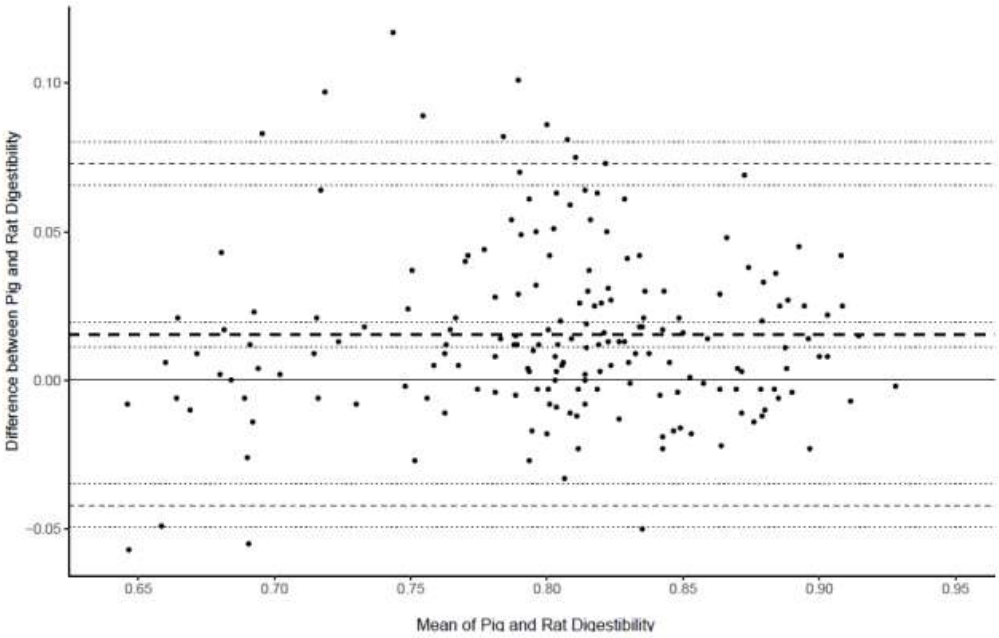


**Figure1.** 0-BA plot of pig and rat digestibility

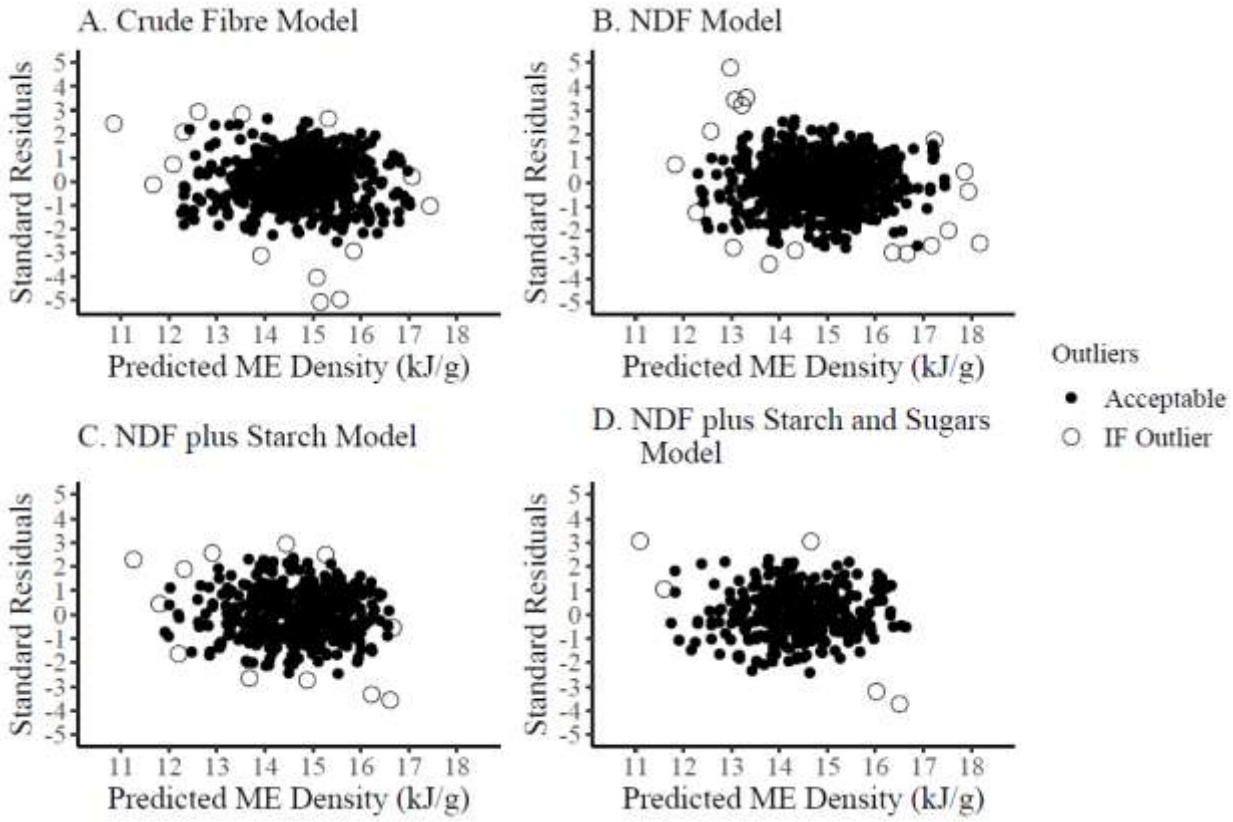


**Figure 2.** Identification of outliers in the ME density data for the four regression models

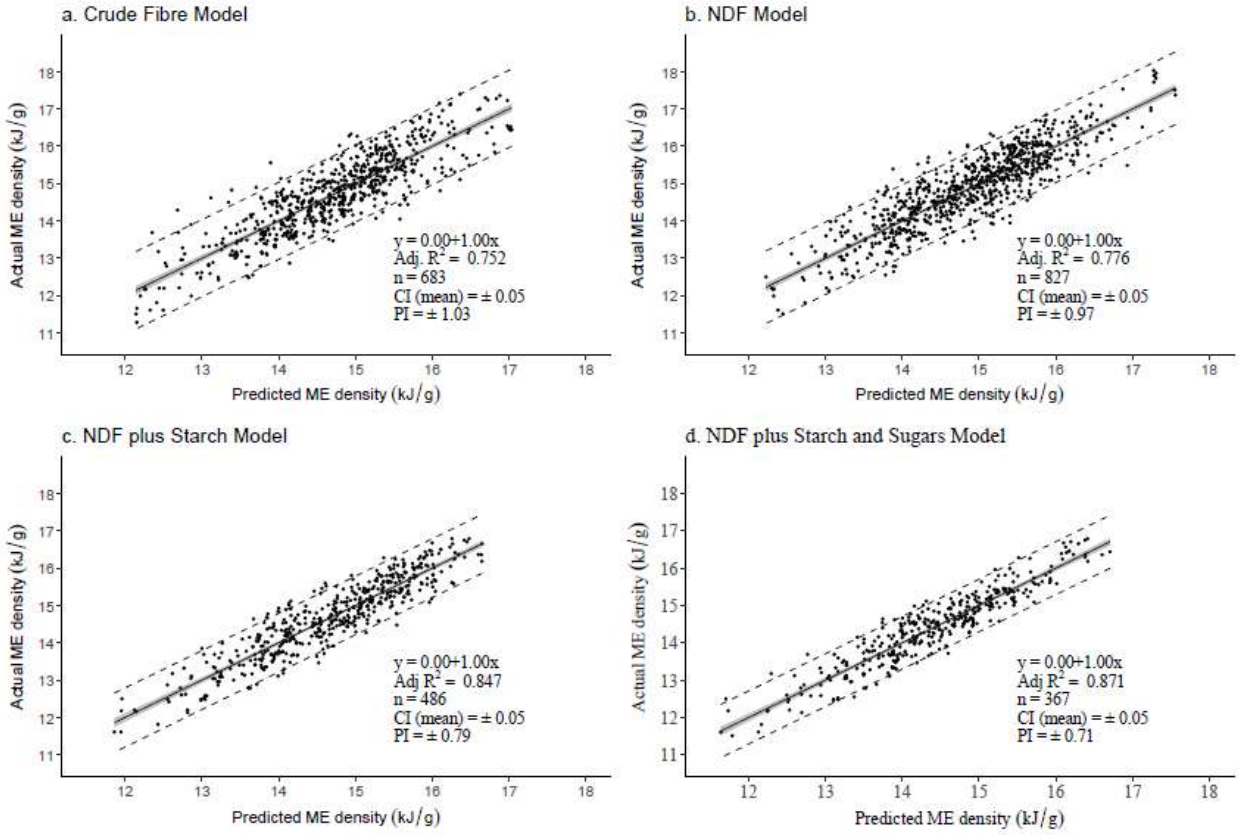


**Figure 3.** Relationship between the measured and predicted ME densities

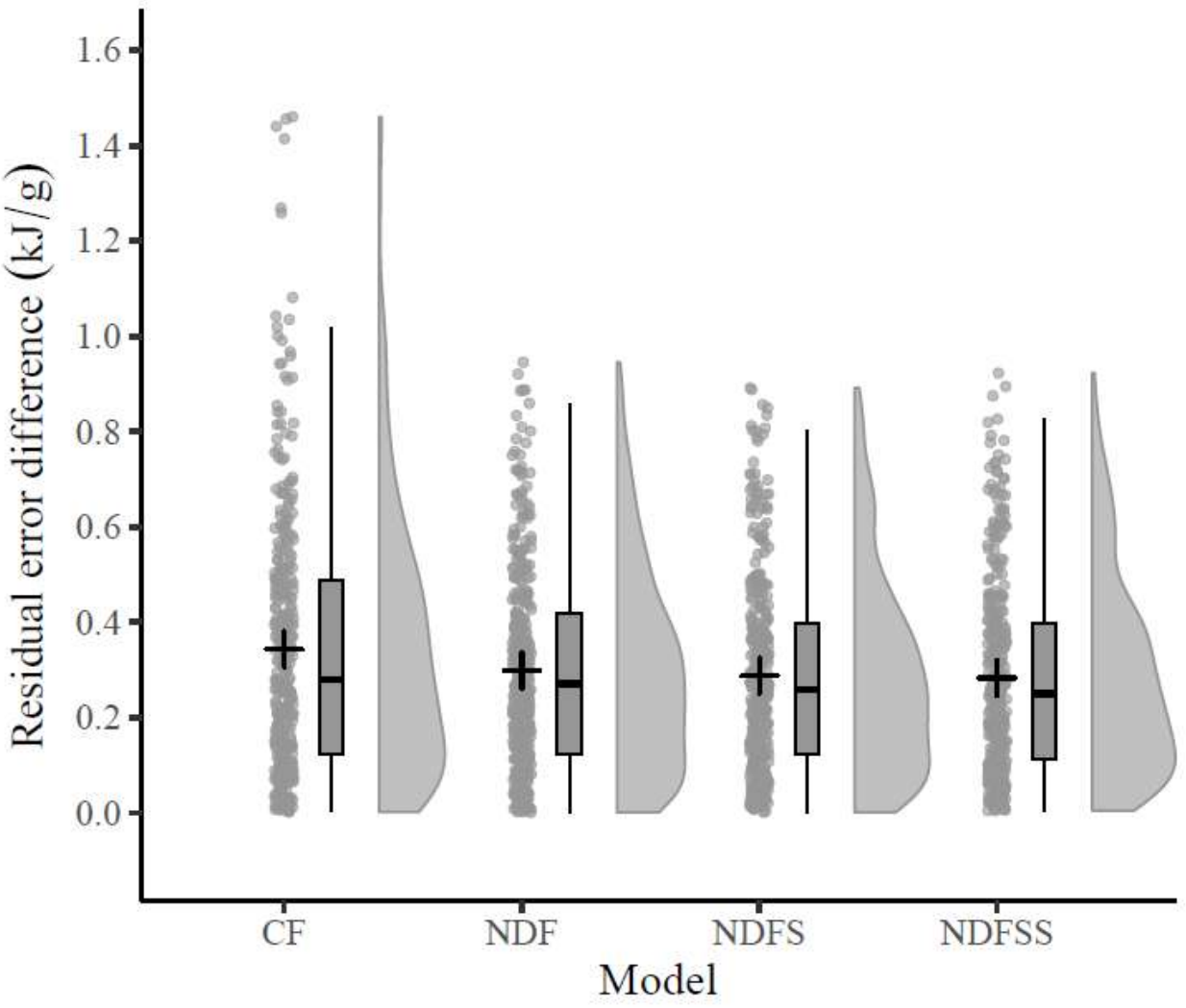


**Figure 4.** Raincloud plot

**Table 1.** Descriptive statistics for four datasets from which outliers have been removed. The chemical analyses are expressed as g per 100 g of dry matter and energy values as kJ/g of dry matter. The abbreviations for the nutrients and energy forms are defined in the abbreviations section.

a. Crude Fibre dataset (N=683)

| | Mean | SD | Median | Min | Max |
|---|---|---|---|---|---|
| CP | 19.10 | 4.98 | 18.88 | 7.83 | 47.20 |
| EE | 4.01 | 2.83 | 2.77 | 0.97 | 19.80 |
| CF | 4.94 | 2.23 | 4.59 | 1.01 | 16.20 |
| NFE | 65.60 | 7.29 | 66.00 | 34.49 | 83.46 |
| Ash | 6.34 | 1.46 | 6.07 | 2.60 | 11.30 |
| GE | 18.44 | 0.71 | 18.30 | 16.36 | 20.98 |
| DE | 15.32 | 1.07 | 15.38 | 11.72 | 18.11 |
| ME | 14.72 | 1.05 | 14.74 | 11.28 | 17.40 |
| ME/DE | 0.96 | 0.01 | 0.96 | 0.91 | 0.99 |
| DE/GE | 0.83 | 0.05 | 0.84 | 0.65 | 0.94 |

b. NDF dataset (N=827)

| | Mean | SD | Median | Min | Max |
|---|---|---|---|---|---|
| CP | 17.89 | 4.08 | 18.44 | 7.17 | 34.66 |
| EE | 4.19 | 2.62 | 3.22 | 0.97 | 15.60 |
| NDF | 16.18 | 5.57 | 15.43 | 4.12 | 46.00 |
| NFC | 55.71 | 8.54 | 55.60 | 20.40 | 78.06 |
| Ash | 6.02 | 1.41 | 5.79 | 3.10 | 10.80 |
| GE | 18.42 | 0.65 | 18.28 | 17.13 | 20.63 |
| DE | 15.45 | 1.04 | 15.53 | 12.16 | 18.57 |
| ME | 14.88 | 1.04 | 14.93 | 11.51 | 18.03 |
| ME/DE | 0.96 | 0.01 | 0.96 | 0.91 | 0.99 |
| DE/GE | 0.84 | 0.05 | 0.84 | 0.65 | 0.95 |

c. NDF plus Starch dataset (N=486)

| | Mean | SD | Median | Min | Max |
|---|---|---|---|---|---|
| CP | 17.87 | 3.81 | 18.34 | 7.17 | 29.00 |
| EE | 3.92 | 2.62 | 2.74 | 0.97 | 11.60 |
| NDF | 15.63 | 4.93 | 14.75 | 4.37 | 31.46 |
| Starch | 47.69 | 10.69 | 47.65 | 2.87 | 76.36 |
| Res1 | 8.48 | 6.24 | 7.71 | -8.10 | 68.73 |
| Ash | 6.40 | 1.45 | 6.11 | 3.30 | 10.83 |
| GE | 18.31 | 0.58 | 18.19 | 17.20 | 20.27 |
| DE | 15.27 | 1.03 | 15.38 | 11.96 | 17.34 |
| ME | 14.68 | 1.02 | 14.71 | 11.61 | 16.79 |
| ME/DE | 0.96 | 0.01 | 0.96 | 0.91 | 0.99 |
| DE/GE | 0.83 | 0.05 | 0.84 | 0.65 | 0.95 |

d. NDF plus Starch and Sugars dataset (N=367)

| | Mean | SD | Median | Min | Max |
|---|---|---|---|---|---|
| CP | 18.24 | 3.36 | 18.49 | 10.05 | 29.00 |
| EE | 3.94 | 2.79 | 2.54 | 0.97 | 11.60 |
| NDF | 16.20 | 4.97 | 15.52 | 4.37 | 29.70 |
| Starch | 46.86 | 9.62 | 47.60 | 10.80 | 68.63 |
| Sugars | 4.65 | 2.61 | 4.17 | 0.20 | 28.24 |
| Res2 | 3.30 | 3.12 | 3.23 | -5.41 | 12.80 |
| Ash | 6.82 | 1.40 | 6.53 | 4.48 | 10.83 |
| GE | 18.29 | 0.60 | 18.16 | 17.20 | 20.27 |
| DE | 15.02 | 1.03 | 15.10 | 11.96 | 17.30 |
| ME | 14.40 | 1.00 | 14.50 | 11.51 | 16.79 |
| ME/DE | 0.96 | 0.01 | 0.96 | 0.91 | 0.98 |
| DE/GE | 0.82 | 0.05 | 0.83 | 0.65 | 0.94 |

**Table 2** Gross and metabolisable energy regressions after removal of outliers, with the Grower/Finisher animals as the reference phase. The values in brackets are the 95% confidence intervals. Unless otherwise stated P<.001. When the coefficients are applied to nutrients expressed as g/100g diet DM, the unit of energy density will be kJ/g DM. Phase AD: the additional energy to be applied to the predicted energy digestibility when used for older animals. The VIF values are identical for the gross and metabolisable energy density regressions.

a. The Crude Fibre dataset

| | Gross Energy Density | | Metabolisable Energy Density | | VIF | |
|---|---|---|---|---|---|---|
| | Residue Model | Ash Model | Residue Model | Ash Model | Residue Model | Ash Model |
| (Intercept) | **2.2083** (0.1508, 4.2658) P=0.035 | **17.4474** (17.3120, 17.5827) | **-16.4063** (-19.5802, -13.2325) | **16.3006** (16.0917, 16.5094) | | |
| CP | **0.2072** (0.1831, 0.2313) | **0.0548** (0.0490, 0.0606) | **0.3894** (0.3522, 0.4265) | **0.0623** (0.0533, 0.0712) | **21.89** | **1.26** |
| EE | **0.3547** (0.3302, 0.3791) | **0.2023** (0.1930, 0.2116) | **0.4686** (0.4309, 0.5064) | **0.1416** (0.1272, 0.1559) | **7.26** | **1.05** |
| CF | **0.1730** (0.1446, 0.2015) | **0.0207** (0.0081, 0.0332) | **0.0649** (0.0210, 0.1088) P=0.004 | **-0.2622** (-0.2815, -0.2429) | **6.14** | **1.19** |
| NFE | **0.1524** (0.1313, 0.1735) | | **0.3271** (0.2945, 0.3596) | | **35.95** | |

| | | | | | | |
|---|---|---|---|---|---|---|
| Ash | | **-0.1524** (-0.1735, -0.1313) | | **-0.3270** (-0.3596, -0.2945) | | **1.44** |
| Phase AD | **0.0064** (-0.0685, 0.0814) P=0.87 | **0.0064** (-0.0685, 0.0814) P=0.87 | **0.2578** (0.1422, 0.3734) | **0.2577** (0.1421, 0.3733) | **1.03** | **1.03** |

b. The NDF dataset

| | Gross Energy Density | | Metabolisable Energy Density | | VIF | |
|---|---|---|---|---|---|---|
| | Residue Model | Ash Model | Residue Model | Ash Model | Residue Model | Ash Model |
| (Intercept) | **1.3475** (-0.1772, 2.8723) P=0.08 | **17.2547** (17.1493, 17.3601) | **-30.8702** (-33.6411, -28.0993) | **17.2736** (17.0820, 17.4652) | | |
| CP | **0.2169** (0.1983, 0.2355) | **0.0579** (0.0526, 0.0631) | **0.5444** (0.5106, 0.5782) | **0.0630** (0.0535, 0.0725) | **16.72** | **1.32** |
| EE | **0.3741** (0.3560, 0.3922) | **0.2151** (0.2078, 0.2223) | **0.6704** (0.6375, 0.7033) | **0.1890** (0.1758, 0.2021) | **6.51** | **1.04** |
| NDF | **0.1704** (0.1538, 0.1870) | **0.0113** (0.0079, 0.0148) | **0.3912** (0.3611, 0.4213) | **-0.0902** (-0.0964, -0.0840) | **24.73** | **1.06** |
| NFC | **0.1591** | | **0.4814** | | **51.11** | |

| | | | | | | |
|---|---|---|---|---|---|---|
| | (0.1435, 0.1746) | | (0.4532, 0.5097) | | | |
| Ash | | **-0.1591** | | **-0.4814** | | **1.39** |
| | | (-0.1746, -0.1435) | | (-0.5097, -0.4532) | | |
| Phase AD | **0.0272** | **0.0272** | **0.3305** | **0.3305** | **1.04** | **1.04** |
| | (-0.0268, 0.0813) P=0.32 | (-0.0268, 0.0813) P=0.32 | (0.2323, 0.4287) | (0.2323, 0.4287) | | |

c. The NDF plus Starch dataset

| | Gross Energy Density | | Metabolisable Energy Density | | VIF | |
|---|---|---|---|---|---|---|
| | Residue Model | Ash Model | Residue Model | Ash Model | Residue Model | Ash Model |
| (Intercept) | **2.6865** | **17.2588** | **-31.6787** | **17.1835** | | |
| | (1.2830, 4.0899) | (16.9903, 17.5273) | (-34.9061, -28.4512) | (16.5661, 17.8009) | | |
| CP | **0.1946** | **0.0488** | **0.5620** | **0.0734** | **17.59** | **1.97** |
| | (0.1774, 0.2117) | (0.0431, 0.0546) | (0.5225, 0.6015) | (0.0602, 0.0866) | | |
| EE | **0.3589** | **0.2132** | **0.6620** | **0.1734** | **8.15** | **1.28** |
| | (0.3419, 0.3759) | (0.2064, 0.2199) | (0.6229, 0.7011) | (0.1579, 0.1889) | | |
| NDF | **0.1604** | **0.0147** | **0.3831** | **-0.1055** | **24.85** | **1.72** |

| | | | | | | |
|---|---|---|---|---|---|---|
| | (0.1447, 0.1762) | (0.0106, 0.0188) | (0.3469, 0.4194) | (-0.1150, -0.0960) | | |
| Starch | **0.1466** | **0.0008** | **0.4930** | **0.0044** | **94.65** | **3.29** |
| | (0.1324, 0.1607) | (-0.0018, 0.0035) P=0.54 | (0.4604, 0.5256) | (-0.0017, 0.0105) P=0.16 | | |
| Res1 | **0.1457** | | **0.4886** | | **34.91** | |
| | (0.1310, 0.1605) | | (0.4547, 0.5226) | | | |
| Ash | | **-0.1457** | | **-0.4886** | | **1.88** |
| | | (-0.1605, -0.1310) | | (-0.5226, -0.4547) | | |
| Phase AD | **0.0175** | **0.0175** | **0.4085** | **0.4085** | **1.11** | **1.11** |
| | (-0.0250, 0.0600) P=0.42 | (-0.0250, 0.0600) P=0.42 | (0.3108, 0.5062) | (0.3108, 0.5062) | | |

d. The NDF plus Starch and Sugars dataset

| | Gross Energy Density | | Metabolisable Energy Density | | VIF | |
|---|---|---|---|---|---|---|
| | Residue Model | Ash Model | Residue Model | Ash Model | Residue Model | Ash Model |
| (Intercept) | **-1.0733** | **18.3393** | **-17.3681** | **13.0380** | | |
| | (-2.6847, 0.5380) P=0.19 | (17.7786, 18.9000) | (-21.4723, -13.2640) | (11.6099, 14.4660) | | |
| CP | **0.2427** | **0.0485** | **0.3961** | **0.0920** | **20.04** | **2.76** |

| | | | | | | |
|---|---|---|---|---|---|---|
| | (0.2232, 0.2621) | (0.0413, 0.0557) | (0.3466, 0.4456) | (0.0737, 0.1104) | | |
| EE | **0.4001** | **0.2059** | **0.4983** | **0.1942** | **13.21** | **1.93** |
| | (0.3811, 0.4191) | (0.1987, 0.2132) | (0.4498, 0.5467) | (0.1757, 0.2127) | | |
| NDF | **0.2000** | **0.0059** | **0.2199** | **-0.0841** | **37.87** | **4.15** |
| | (0.1820, 0.2181) | (-0.0001, 0.0119) P=0.053 | (0.1740, 0.2659) | (-0.0993, -0.0689) | | |
| Starch | **0.1844** | **-0.0097** | **0.3479** | **0.0438** | **115.03** | **13.40** |
| | (0.1682, 0.2006) | (-0.0153, -0.0042) P=0.001 | (0.3065, 0.3893) | (0.0297, 0.0579) | | |
| Sugars | **0.1809** | **-0.0132** | **0.3429** | **0.0388** | **11.64** | **2.44** |
| | (0.1619, 0.1999) | (-0.0219, -0.0045) P=0.003 | (0.2945, 0.3913) | (0.0166, 0.0610) P=0.001 | | |
| Res2 | **0.1941** | | **0.3041** | | **17.00** | |
| | (0.1749, 0.2134) | | (0.2551, 0.3531) | | | |
| Ash | | **-0.1941** | | **-0.3041** | | **3.42** |
| | | (-0.2134, -0.1749) | | (-0.3531, -0.2551) | | |
| Phase AD | **0.0067** | **0.0067** | **0.4750** | **0.4750** | **1.18** | **1.18** |
| | (-0.0300, 0.0433) P=0.72 | (-0.0300, 0.0433) P=0.72 | (0.3817, 0.5684) | (0.3817, 0.5684) | | |

**Table 3.** Calculated gross and metabolisable energy density of nutrients.

| Nutrient | GE Density (kJ/g) | | Median Apparent Digestibility | DE Density (kJ/g) | Fibre Gaseous loss (% DE) | After Fibre Gaseous loss | ME:DE Ratio | ME Density (kJ/g) | |
|---|---|---|---|---|---|---|---|---|---|
| | Publication-based | Our estimate | | | | | | Calculated | Our estimate from regressions |
| Crude protein | 23.71 | 23.05 | 0.80 | 18.97 | | | 0.83 | 15.74 | 18.79 |
| Ether extract | 39.22 | 38.62 | 0.75 | 29.41 | | | 1.00 | 29.41 | 30.55 |
| Crude fibre | 19.83 | 20.14 | 0.40 | 7.93 | 5 | 7.54 | 1.00 | 7.54 | -15.73 |
| Neutral Detergent Fibre | 17.41 | 18.79 | 0.60 | 10.45 | 5 | 9.92 | 1.00 | 9.92 | 3.77 |
| Nitrogen-free extractives | 17.39 | 17.50 | 0.90 | 15.65 | | | 1.00 | 15.65 | 15.96 |
| Non-fibrous Carbohydrates | 17.32 | 17.29 | 1.00 | 17.32 | | | 1.00 | 17.32 | 16.68 |
| Starch | 17.46 | 17.37 | 1.00 | 17.46 | | | 1.00 | 17.46 | 17.54 |
| Sugars | 16.14 | 16.89 | 1.00 | 16.14 | | | 1.00 | 16.14 | 15.28 |

1. Our estimates of GE and ME density are taken from the non-intercept models, using mean values where possible. The ME values were from the grower/finisher data only.
2. The basis for the data taken directly or calculated from publications is provided in the supplementary data.
3. The values are estimated for typical diets with a digestibility greater than about 0.7; the energy density values for protein and fat are likely to be substantially lower in high-fibre diets.
4. We have rounded our estimates on nutrient digestibility to the nearest 0.05 to avoid suggesting that the values are more than informed estimates from the literature. They could vary by 0.05 to 0.1 units.

**Table 4** Goodness of fit estimates for the gross and metabolisable energy density intercept-based regressions of the combined GF and AD data.

a. Gross Energy

| Parameter | CF Model | NDF Model | NDF plus Starch Model | NDF plus Starch and Sugars Model |
|---|---|---|---|---|
| N | 683 | 827 | 486 | 367 |
| Adj. $R^2$ | 0.769 | 0.827 | 0.908 | 0.945 |
| Pred. $R^2$ | 0.766 | 0.825 | 0.907 | 0.943 |
| RMSE | 0.340 | 0.271 | 0.173 | 0.140 |
| RSD/RSE | 0.341 | 0.272 | 0.175 | 0.142 |
| MAE | 0.215 | 0.196 | 0.133 | 0.110 |
| (95% CI) | (0.196, 0.235) | (0.183, 0.209) | (0.123, 0.143) | (0.102, 0.119) |
| MAPE (%) | 1.167 | 1.061 | 0.729 | 0.603 |
| AIC | 477.061 | 200.649 | -307.751 | -382.911 |

b. Metabolizable Energy

| Parameter | CF Model | NDF Model | NDF plus Starch Model | NDF plus Starch and Sugars Model |
|---|---|---|---|---|
| N | 683 | 827 | 486 | 367 |
| Adj. $R^2$ | 0.751 | 0.775 | 0.845 | 0.869 |
| Pred. $R^2$ | 0.747 | 0.772 | 0.842 | 0.865 |
| RMSE | 0.524 | 0.492 | 0.399 | 0.357 |
| RSD/RSE | 0.526 | 0.494 | 0.402 | 0.361 |
| MAE | 0.422 | 0.391 | 0.324 | 0.289 |
| (95% CI) | (0.399, 0.446) | (0.370, 0.411) | (0.303, 0.345) | (0.268, 0.311) |
| MAPE (%) | 2.880 | 2.652 | 2.224 | 2.031 |
| AIC | 1069.173 | 1188.665 | 501.709 | 303.331 |

**Table 5** No intercept regressions. The data for the grower/finisher and adult animals are combined to estimate the gross energy density coefficients but reported separately for the metabolisable energy density coefficients (see text for explanation).

a. Gross Energy Density

| | CF Model | NDF Model | NDFS Model | NDFSS Model |
|---|---|---|---|---|
| CP | **0.2326** | **0.2328** | **0.2267** | **0.2300** |
| | (0.2284, 0.2368) | (0.2291, 0.2366) | (0.2231, 0.2303) | (0.2259, 0.2341) |
| EE | **0.3790** | **0.3886** | **0.3891** | **0.3879** |
| | (0.3700, 0.3881) | (0.3816, 0.3957) | (0.3830, 0.3952) | (0.3824, 0.3934) |
| CF | **0.2014** | | | |
| | (0.1905, 0.2122) | | | |
| NFE | **0.1750** | | | |
| | (0.1738, 0.1763) | | | |
| NDF | | **0.1851** | **0.1904** | **0.1882** |
| | | (0.1821, 0.1881) | (0.1875, 0.1933) | (0.1857, 0.1908) |
| NFC | | **0.1729** | | |
| | | (0.1718, 0.1739) | | |
| Starch | | | **0.1737** | **0.1736** |
| | | | (0.1728, 0.1746) | (0.1727, 0.1746) |
| Res1 | | | **0.1733** | |
| | | | (0.1707, 0.1759) | |
| Sugars | | | | **0.1689** |
| | | | | (0.1625, 0.1754) |
| Res2 | | | | **0.1815** |
| | | | | (0.1766, 0.1864) |

1. Values in parentheses are the 95% confidence intervals.
2. Unless otherwise shown, the p values are < 0.001.

b. Metabolisable Energy Density

| | Grower/Finisher data | | | | Adult data | | | |
|---|---|---|---|---|---|---|---|---|
| | CF Model | NDF Model | NDFS Model | NDFSS Model | CF Model | NDF Model | NDFS Model | NDFSS Model |
| CP | **0.2014** (0.1941, 0.2087) | **0.1769** (0.1674, 0.1864) | **0.1830** (0.1710, 0.1950) | **0.1901** (0.1773, 0.2029) | **0.1763** (0.1516, 0.2010) | **0.1666** (0.1498, 0.1835) | **0.1900** (0.1681, 0.2118) | **0.1995** (0.1751, 0.2239) |
| EE | **0.2889** (0.2724, 0.3053) | **0.3315** (0.3134, 0.3496) | **0.2997** (0.2782, 0.3211) | **0.3019** (0.2846, 0.3193) | **0.2843** (0.2516, 0.3170) | **0.3476** (0.3203, 0.3749) | **0.3141** (0.2855, 0.3428) | **0.3062** (0.2794, 0.3330) |
| CF | **-0.1573** (-0.1770, -0.1376) | | | | **-0.0505** (-0.0894, -0.0115) P=0.012 | | | |
| NFE | **0.1596** (0.1573, 0.1618) | | | | **0.1613** (0.1558, 0.1668) | | | |
| NDF | | **0.0618** (0.0541, 0.0695) | **0.0291** (0.0188, 0.0394) | **0.0221** (0.0138, 0.0303) | | **0.0608** (0.0489, 0.0726) | **0.0602** (0.0461, 0.0742) | **0.0530** (0.0396, 0.0664) |
| NFC | | **0.1668** (0.1642, | | | | **0.1737** (0.1695, | | |

| | Grower/Finisher data | | | | Adult data | | | |
|---|---|---|---|---|---|---|---|---|
| | CF Model | NDF Model | NDFS Model | NDFSS Model | CF Model | NDF Model | NDFS Model | NDFSS Model |
| | | 0.1694) | | | | 0.1779) | | |
| Starch | | | **0.1747** (0.1716, 0.1778) | **0.1760** (0.1729, 0.1792) | | | **0.1696** (0.1647, 0.1745) | **0.1701** (0.1655, 0.1746) |
| Res1 | | | **0.1621** (0.1533, 0.1709) | | | | **0.1538** (0.1295, 0.1780) | |
| Sugars | | | | **0.1528** (0.1341, 0.1715) | | | | **0.1374** (0.0647, 0.2101) |
| Res2 | | | | **0.0948** (0.0792, 0.1104) | | | | **0.1552** (0.1270, 0.1833) |

1. Values in parentheses are the 95% confidence intervals.
2. Unless otherwise shown, the P values are < 0.001.

**Table 6** Goodness of fit estimates for the gross and metabolisable energy density no-intercept regressions of the grower/finisher and adult data in Table 5

| | Gross Energy Density: GF and AD combined | | | | Metabolisable Energy Density: Grower/Finisher data only | | | | Metabolisable Energy Density: Adult data only | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Parameter | CF Model | NDF Model | NDFS Model | NDFSS Model | CF Model | NDF Model | NDFS Model | NDFSS Model | CF Model | NDF Model | NDFS Model | NDFSS Model |
| N | 683 | 827 | 486 | 367 | 587 | 709 | 397 | 276 | 96 | 118 | 89 | 91 |
| RMSE | 0.341 | 0.272 | 0.176 | 0.140 | 0.576 | 0.651 | 0.562 | 0.392 | 0.413 | 0.369 | 0.321 | 0.307 |
| RSD/RSE | 0.342 | 0.272 | 0.177 | 0.142 | 0.578 | 0.653 | 0.566 | 0.396 | 0.422 | 0.376 | 0.330 | 0.318 |
| MAE | 0.215 | 0.197 | 0.133 | 0.111 | 0.461 | 0.524 | 0.452 | 0.314 | 0.332 | 0.284 | 0.239 | 0.233 |
| (95% CI) | (0.195, 0.235) | (0.184, 0.210) | (0.123, 0.143) | (0.102, 0.120) | (0.433, 0.489) | (0.495, 0.552) | (0.419, 0.485) | (0.286, 0.342) | (0.282, 0.382) | (0.241, 0.327) | (0.194, 0.284) | (0.191, 0.275) |
| MAPE (%) | 1.164 | 1.069 | 0.727 | 0.607 | 3.155 | 3.584 | 3.122 | 2.224 | 2.268 | 1.933 | 1.651 | 1.616 |
| AIC | 477.6 | 200.9 | -296.9 | -385.0 | 1029.2 | 1413.0 | 681.1 | 280.2 | 112.6 | 109.9 | 62.3 | 57.4 |

1. Values for $R^2$ are inappropriate for no-intercept regressions and are excluded.

**Table 7.** Comparison of goodness of fit measures of the four datasets of common grower/finisher and adult records

a. Gross Energy

| Parameter | CF Model | NDF Model | NDFS Model | NDFSS Model |
|---|---|---|---|---|
| N | 359 | 359 | 359 | 359 |
| Adj. $R^2$ | 0.943 | 0.944 | 0.944 | 0.945 |
| Pred. $R^2$ | 0.942 | 0.943 | 0.943 | 0.943 |
| RMSE | 0.144 | 0.143 | 0.142 | 0.141 |
| RSD/RSE | 0.145 | 0.144 | 0.143 | 0.142 |
| MAE | 0.114 | 0.112 | 0.111 | 0.111 |
| (95% CI) | (0.105, 0.123) | (0.103, 0.121) | (0.101, 0.120) | (0.102, 0.120) |
| MAPE (%) | 0.621 | 0.610 | 0.605 | 0.606 |
| AIC | -359.502 | -365.757 | -366.330 | -371.806 |

b. Metabolizable Energy

| Parameter | CF Model | NDF Model | NDFS Model | NDFSS Model |
|---|---|---|---|---|
| N | 359 | 359 | 359 | 359 |
| Adj. $R^2$ | 0.784 | 0.851 | 0.861 | 0.865 |
| Pred. $R^2$ | 0.779 | 0.848 | 0.857 | 0.861 |
| RMSE | 0.442 | 0.367 | 0.354 | 0.348 |
| RSD/RSE | 0.446 | 0.370 | 0.357 | 0.352 |
| MAE | 0.343 | 0.299 | 0.288 | 0.283 |
| (95% CI) | (0.314, 0.372) | (0.277, 0.321) | (0.266, 0.309) | (0.261, 0.304) |
| MAPE (%) | 2.399 | 2.087 | 2.008 | 1.973 |
| AIC | 446.726 | 313.255 | 288.971 | 279.836 |

**Table 8** The probable range of goodness of fit measures for ME density on future datasets with similar characteristics to the diets in this study, determined by a 10-fold cross-validation analysis repeated 10 times. Values in parenthesis are the 95% CI.

| GOF Indices | CF | NDF | NDFS | NDFSS |
|---|---|---|---|---|
| $R^2$ | 0.751 *(0.740, 0.762)* | 0.775 *(0.767, 0.783)* | 0.844 *(0.838, 0.851)* | 0.867 *(0.859, 0.875)* |
| Predicted $R^2$ | 0.747 | 0.772 | 0.842 | 0.865 |
| Adjusted $R^2$ | 0.747 *(0.736, 0.758)* | 0.772 *(0.764, 0.780)* | 0.841 *(0.834, 0.848)* | 0.863 *(0.855, 0.872)* |
| RMSE | 0.528 *(0.519, 0.536)* | 0.495 *(0.488, 0.502)* | 0.404 *(0.396, 0.411)* | 0.364 *(0.356, 0.372)* |
| RSD | 0.527 *(0.518, 0.535)* | 0.495 *(0.488, 0.502)* | 0.406 *(0.398, 0.413)* | 0.363 *(0.355, 0.371)* |
| MAE | 0.427 *(0.419, 0.434)* | 0.394 *(0.388, 0.400)* | 0.329 *(0.322, 0.335)* | 0.296 *(0.289, 0.303)* |

1. The Predicted $R^2$ values from Table 4b have been included for comparison.

**Table 9.** Published equations for the prediction of ME density from analytical components of diets standardised as g/100g for nutrients and kJ/g for energy. The regression equations have been applied to the average nutrient values for the 271 grower/finisher records from the common dataset to give a predicted ME density. The average measured ME density of the 271 records was 14.34 kJ/g. For consistency, the predicted ME densities and GOF indices have been rounded to two decimal places. See footnote 1 for the definition of the residues (Res1 to 4).

**Intercept Models**

| Source | # Diets | Model | Equation | Predict. ME | $R^2$ | RSD | RMSE |
|---|---|---|---|---|---|---|---|
| A | 21 Ingred + 16 diets | Equation 35 | -20.0623 + 0.4163 CP + 0.6054 AEE + 0.3674 NFE | 14.06 | 0.90 | 0.89 | |
| | | Our regression | -19.8385 + 0.4147 CP + 0.5858 EE – 0.0048 CF + 0.3677 NFE | 14.17 | 0.88 | | 0.95 |
| B | 24 Diets | CF & NFE v6 | -8.529 + 0.255 CP + 0.338 EE + 0.036 CF + 0.266 Starch + 0.293 Sugars + 0.200 Res3 | 14.31 | 0.90 | 0.22 | |
| C | 321 Diets | Table XV | 5.412 + 0.143 CP + 0.194 EE - 0.245 CF + 0.106 NFE | 14.46 | 0.77 | 0.72 | |
| | 50 Diets | Table XVI | -34.627 + 0.632 CP + 0.624 EE + 0.109 CF + 0.516 NFE | 13.83 | 0.96 | 0.32 | |
| D | 114 Diets | Equation 34 | 17.4391 + 0.05858 CP + 0.1715 EE - 0.2552 CF - 0.5146 Ash | 14.30 | 0.88 | 0.34 | |
| | | Equation 35 | 17.5477 + 0.0418 CP + 0.1715 EE - 0.1464 NDF - 0.3849 Ash | 14.08 | 0.92 | 0.29 | |
| E | 77 Diets | Equation 26 | 17.64 + 0.039 CP + 0.151 EE - 0.135 NDF - 0.323 Ash | 14.64 | 0.91 | 0.30 | |
| F | 587 GF Diets | CF & NFE | -16.4063 + 0.3894 CP + 0.4686 EE + 0.0649 CF + 0.3271 NFE | 14.38 | 0.75 | 0.53 | 0.52 |
| | 587 GF Diets | CF & Ash | 16.3006 + 0.0623 CP + 0.1416 EE - 0.2622 CF - 0.3270 Ash | 14.38 | | | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | 709 GF Diets | NDF & NFC | -30.8702 + 0.5444 CP + 0.6704 EE + 0.3912 NDF + 0.4814 NFC | 14.47 | 0.78 | 0.49 | 0.49 |
| | 709 GF Diets | NDF & Ash | 17.2736 + 0.0630 CP + 0.1890 EE - 0.0902 NDF -0.4814 Ash | 14.47 | | | |
| | 397 GF Diets | NDFS | -31.6787 + 0.5620 CP + 0.6620 EE + 0.3831 NDF + 0.4930 Starch + 0.4886 Res1 | 14.43 | 0.85 | 0.40 | 0.40 |
| | 397 GF Diets | NDFS & Ash | 17.1835 + 0.0734 CP + 0.1734 EE -0.1055 NDF +0.0044 Starch - 0.4886 Ash | 14.43 | | | |
| | 276 GF Diets | NDFSS & NDF | -17.3681 - 0.3961 CP + 0.4983 EE + 0.2199 NDF + 0.3479 Starch + 0.3429 Sugars + 0.3041 Res2 | 14.34 | 0.87 | 0.36 | 0.36 |
| | 276 GF Diets | NDFSS & Ash | 13.0380 + 0.0920 CP + 0.1942EE - 0.0841 NDF + 0.0438 Starch + 0.0388 Sugars - 0.3041 Ash | 14.33 | | | |

**No-intercept Models**

| Source | # Diets | Model | Equation | Predict. ME | $R^2$ | RSD | RMSE |
|---|---|---|---|---|---|---|---|
| A | 21 Ingred + 16 diets | Our regression | 0.1735 CP + 0.3897 EE - 0.2450 CF + 0.1685 NFE | 14.49 | | | 1.35 |
| C | 321 diets | Table XV | 0.203 CP+ 0.252EE - 0.178 CF + 0.162 NFE | 14.40 | | 0.32 | |
| | 50 diets | Table XVI | 0.236 CP + 0.260 EE - 0.272 CF + 0.159 NFE | 14.36 | | 0.36 | |
| E | 77 Diets | Equation 26 (G only) | 0.201 CP + 0.318 EE + 0.026 NDF + 0.171 Starch + 0.165 Res1# | 14.77 | | 0.34 | |
| F | 587 GF Diets | CF Model | 0.2014 CP + 0.2889 EE - 0.1573 CF + 0.1596 NFE | 14.48 | | 0.58 | 0.58 |
| | 709 GF Diets | NDF Model | 0.1769 CP + 0.3315 EE + 0.0618 NDF + 0.1668 NFC | 14.75 | | 0.65 | 0.65 |
| | 397 GF Diets | NDFS Model | 0.1830 CP + 0.2997 EE + 0.0291 NDF + 0.1747 Starch + 0.1621 Res1# | 14.55 | | 0.57 | 0.56 |
| | 276 GF Diets | NDFSS Model | 0.1901 CP + 0.3019 EE + 0.0221 NDF + 0.1760 Starch + 0.1528 Sugars + 0.0948 Res2# | 14.34 | | 0.40 | 0.39 |
| G | 48 Diets | Table 2 (ME from DE less urine) | 0.226 CP + 0.319 EE - 0.129 CF + 0.166 Starch + 0.184 Sugars + 0.097 Res3# | 14.74 | | 0.37 | |
| | | Table 2 | 0.223 CP + 0.341 EE -0.109 CF + 0.170 Starch + 0.168 Sugars + | 14.66 | | 0.28 | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | (calculated $ME_{BFS}$) | 0.074 Res3# | | | | |
| H | Piglets & GF 290 Diets | Model B1 | 0.20930 CP + 0.32846 EE - 0.23898 CF + 0.16458 Starch +0.21068 Sugars + 0.13962 Res3# | 14.55 | | | 0.24 |
| | Piglets & GF 290 Diets | Model C1 | 0.21503 CP + 0.32497 EE - 0.21071 CF + 0.16309 Starch + 0.14701 Res4# | 14.52 | | | 0.25 |
| | Piglets & GF 290 Diets | Model D1 | 0.21041 CP + 0.33050 EE - 0.25123 CF + 0.16164 NFE | 14.45 | | | 0.26 |
| | Piglets 92 Diets | Model B2 | 0.19395 CP + 0.43018 EE - 0.14847 CF + 0.16182 Starch + 0.15015 Sugars + 0.14131 Res3# | 14.76 | | | 0.23 |
| | Piglets 92 Diets | Model C2 | 0.19329 CP + 0.43655 EE - 0.14559 CF + 0.16131 Starch + 0.14313 Res4# | 14.76 | | | 0.22 |
| | Piglets 92 Diets | Model D2 | 0.19586 CP + 0.42760 EE - 0.19627 CF + 0.15791 NFE | 14.63 | | | 0.22 |
| | G/F 198 Diets | Model B3 | 0.20984 CP + 0.31895 EE - 0.23710 CF + 0.16556 Starch + 0.21359 Sugars + 0.13495 Res3# | 14.53 | | | 0.24 |
| | G/F 198 Diets | Model C3 | 0.21984 CP + 0.30506 EE - 0.20134 CF + 0.16413 Starch + 0.13961 Res4# | 14.48 | | | 0.25 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | G/F 198 Diets | Model D3 | 0.21197 CP + 0.31849 EE - 0.26291 CF + 0.16224 NFE | 14.41 | | | 0.27 |
| I | 524 Diets | C1 Piglets & GF | 0.2039 CP + 0.3702 EE -0.1775 CF + 0.1648 Starch + 0.1412 Res 4[#] | 14.64 | | | 0.32 |

1. # Organic residue definitions:
   Res1: 100-(CP+EE+NDF+Starch+Ash);

   Res2: 100-(CP+EE+NDF+Starch+Sugars+Ash);

   Res3: 100-(CP+EE+CF+Starch+Sugars+Ash);

   Res4: 100-(CP+EE+CF+Starch+Ash)

2. Source key: A - Morgan et al., 1975 II (see 3 below), B - Eeckhout and Moermans, 1981, C - Just et al., 1984, D - Noblet & Perez 1993, E - Le Goff & Noblet, 2001, F – Current paper, G - Kirchgessner & Roth, 1983, H - Bulang and Rodehutscord 2009, I - Grümpel-Schlüter et al., 2021.
3. Mean analytical values (%, g/100g) for the 271 G/F records in the common dataset:
   CP 18.50, EE 4.10, CF 5.29, NFE 65.17, NDF 15.55, NFC 54.91, Starch 46.15, Sugars 4.92, Res1 8.77, Res2 3.84, Res3 14.09, Res4 19.02, Ash 6.94.
4. The Morgan data combines ingredient and diet data. Their original Eq 35 uses acid hydrolysis ether extract as a coefficient. We have used the ether extract value, which introduces a small error in the predicted ME value. We carried out a multiple regression analysis of Morgan's data to estimate the coefficients of intercept and no-intercept models that included CP, EE, CF and NFE. Our regression based on CP, EE, and NFE gave coefficients nearly identical to those of Morgan's Eq 35 despite substituting EE for AEE.
5. For our intercept-based regressions (F), the GOF values are from the combined GF and Adult data.
6. For reasons given in the text, $R^2$ values are not relevant for no-intercept models and are excluded even when given in a publication.