A Novel Approach to Calculating Energy Density from Food Images Reduces Analysis Time and Cost

Zhaoxing Pan¹, Tyson Marden², Archana Mande², Janine Higgins²,³*

¹Research Institute, Children’s Hospital Colorado, Aurora, CO, USA
²Colorado Clinical & Translational Sciences Institute, University of Colorado Anschutz Medical Campus, Aurora, CO, USA
³Department of Pediatrics, School of Medicine, University of Colorado Anschutz Medical Campus, Aurora, CO, USA
Email: *Janine.Higgins@childrenscolorado.org

Abstract

Traditional methods of self-reported food intake are characterized by limitations such as underreporting, high participant burden, and high cost. With the development of automated devices to capture food images and monitor food intake, an accurate and efficient method to estimate energy intake is needed. This study aimed to develop an accurate and time efficient method for estimating energy intake from food images by defining a simple and less burdensome way of estimating energy density (ED). Four experimental methods, exchange, food score-long, food score-short, and meal, were developed to estimate ED based on nutrient composition, water content, and relative proportion of foods in images, using different approaches. Three trained nutritionists analyzed 29 food images for ED using each method. All four experimental methods overestimated ED compared to the FVM but the meal method exhibited the closest agreement, lowest variance for ED, and significantly decreased analysis time by an average of 53 s/meal (p = 0.03). The meal method was used for full-scale validation by analyzing 213 food images against weighed food records. The meal method reduced analysis time by 69% (120 s; p ≤ 0.0001) and over-estimated ED by an average of 1.56 ± 3.17 J/g (p < 0.0001) compared to the FVM and 1.67 ± 3.09 J/g (p < 0.0001) compared to the WFR. The meal method is a novel and quick approach to calculate ED from dietary intake images.

Keywords

Energy Density, Self-Report, Dietary Intake, Eating Behavior, Energy Intake,
1. Introduction

Food intake resulting in over- or under-nutrition is linked to many health problems including obesity, Type 2 Diabetes, cardiovascular disease, and failure to thrive [1] [2] [3]. Collection of dietary information using traditional methods is a tedious process involving self-report. Estimates indicate that participants routinely under-report energy intake by 20% - 50% [1] [4] [5] thereby reducing the accuracy of the information collected. Other limitations of traditional methods include high participant burden, which can change habitual eating behavior, and high cost [1] [2] [3]. Therefore, novel methods that would be accurate, easy to implement, and faster/cheaper than traditional methods are urgently needed.

One of the earliest automated devices developed to measure food intake was the Universal Eating Monitor [6] that permits covert weighing of a participant’s plate every 3 seconds. This method was novel though not compatible with free-living situations. Recently, several methods of automated dietary intake assessment have been reported which improve accuracy, reduce or eliminate self-report, and decrease participant burden [7]-[14]. Most of these novel methods utilize technology ranging from mobile phones to sophisticated wearable sensors that capture eating events and food images [7] [11] [12]. However, most automated methods are expensive, and continue to rely on manual analysis of food images, which is time consuming and further increases cost.

There is a growing body of literature focused on measuring the microstructure of food intake, which includes factors such as eating episode duration, duration of actual ingestion, the number of eating events, rate of ingestion, chewing frequency, chewing efficiency, and bite size [13]-[18]. Automated devices that facilitate the capture of meal microstructure and provide a better understanding of eating behaviors could provide additional benefit for those aiming to reduce energy intake and/or provide more effective self-assessment and feedback tools for those on a restricted diet. Analysis of food images from digital devices to accurately measure energy intake is still mostly manual but is an area of ongoing research. Food image analysis by nutritionists reduces participant burden by shifting responsibility for portion size estimation to trained personnel. It is argued that the increased cost of this trained staff time is offset by reduced participant burden and increased accuracy [7] [19]. Methods facilitating fully automated image analysis need complex algorithms, encounter food recognition issues, and often suffer from the inability to distinguish between similar ingredients or differing preparation styles [11] [12] [20]. A method involving accelerated manual analysis of food images using a standardized procedure by trained staff could be a viable solution to address some of these challenges [7] [19].

This study was conducted to develop an accurate and cost-efficient method for estimating energy intake from food images in free-living populations. We
hypothesize that an accelerated method of visually analyzing food images will be as accurate as WFR and more time-efficient than the FVM, thereby lowering the overall cost for estimation of energy intake from photographic food records.

2. Materials and Methods

Energy density (ED) refers to the amount of energy in a given weight of food (J/g). The water content (W) of food is a primary determinant of ED because it adds weight but no energy [21], whereas fat (3.77 J/g) increases the ED of a food to a greater extent than either carbohydrate or protein (1.67 J/g) [22] [23]. The United States Department of Agriculture (USDA) database [24] was used to assign the W of foods on a per g basis (scored as 0.01 - 1.00 for 1% - 100% water, respectively). W was included in calculations for all experimental methods tested (SI Table 1 & Table 2). In the exchange and food score methods, W of the entire meal was calculated from the combined individual water contents and relative proportion of each food in an image:

\[
\text{Total } W = (W_1 \times \text{Proportion of Food } 1) + (W_2 \times \text{Proportion of Food } 2) + \ldots + (W_n \times \text{Proportion of Food } n)
\]

where \(W\) is water content from the USDA database (1)

In this study, four experimental methods were developed to analyze ED from food images. All methods derive ED based on the nutrient composition, relative food proportions, and W. However, each method follows a different approach to incorporate these factors to yield ED.

2.1. Weighed Food Records

Weighed Food Records (WFR) are considered the “gold standard” of individual

<p>| Table 1. Accuracy and time for ED estimation from food images in Phase 1. |
|-------------------|-------------------|-------------------|-------------------|</p>
<table>
<thead>
<tr>
<th><strong>Method</strong></th>
<th><strong>Mean difference from Full Visual Method (kJ)</strong></th>
<th><strong>p-value (Mean difference from Full Visual Method)</strong></th>
<th><strong>Time per image (s)</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Full Visual Method</td>
<td>157 ± 48</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exchange</td>
<td>1.8 ± 0.50</td>
<td>0.0013</td>
<td>104 ± 36*</td>
</tr>
<tr>
<td>Food Score Long</td>
<td>2.5 ± 0.56</td>
<td>0.0002</td>
<td>117 ± 49*</td>
</tr>
<tr>
<td>Food Score Short</td>
<td>2.2 ± 0.85</td>
<td>0.0134</td>
<td>116 ± 48*</td>
</tr>
<tr>
<td>Meal</td>
<td>1.6 ± 0.46</td>
<td>0.0019</td>
<td>68 ± 25*</td>
</tr>
</tbody>
</table>

Accuracy data are represented as mean +/- SEM; *p < 0.0001 for difference from FVM.

<p>| Table 2. Accuracy for ED estimation. |
|-------------------|-------------------|-------------------|</p>
<table>
<thead>
<tr>
<th><strong>Method</strong></th>
<th><strong>Mean ED (J/g)</strong></th>
<th><strong>p-value for difference from WFR</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>WFR</td>
<td>6.68 ± 0.004</td>
<td>-</td>
</tr>
<tr>
<td>Full Visual</td>
<td>6.64 ± 0.02</td>
<td>0.538</td>
</tr>
<tr>
<td>Meal</td>
<td>8.04 ± 0.02</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

Accuracy data are represented as mean ± SEM; Meal Method is average of all three users.
quantitative dietary assessment methods [25]. WFR require the respondent or recorder to weigh all foods and beverages at the time of consumption. Any plate waste is also be recorded.

Though no dietary assessment methodology can completely prevent measurement error, WFR are considered the most accurate method when it comes to quantifying food intake, since each food is weighed, eliminating issues associated with portion size estimation and recall bias. As a result, WFR are used as the gold standard reference method for validation of other dietary assessment methods [26] [27]. However, administration of WFR can be difficult in many populations and environments such as school age children and work places. Significant training of the recorder is required to minimize errors in data collection and WFR are intrusive and, so, can disrupt participant eating behavior [28].

For this study, data was used from a previous protocol [19] where participants consumed a weighed, metabolic diet for 3 days and returned any uneaten items for weigh back the next day. Each food item was weighed in and out separately. A total of 213 meals were analyzed using WFR [19].

### 2.2. Full Visual Method

Photographic food records used to capture free-living food intake utilize manual interpretation of before and after pictures to estimate food intake at a given meal by trained nutritionists. In a previous study, this method was found to be as accurate as and more convenient for participants than traditional diet diaries [12].

The FVM involved visual estimation of volume of food ingested from pre- and post-meal images. Serving sizes were estimated relative to the plate or package size and the total field of view. Data from every individual food consumed was entered into Nutrient Data Systems for Research (NDS-R; University of Minnesota) software [24].

\[
ED_{\text{full visual method}(j,g)} = \frac{\text{Total Ingested Energy (J)}}{\text{Total Meal Weight (g)}}
\]  

### 2.3. Experimental Methods

In the exchange and food score methods, the proportion of each food item was visually estimated based on the volume of a food in relation to the total volume of all foods in the image. This process was conducted on pre-meal images only; no estimation of actual intake volume was estimated using post-meal images as the purpose was to calculate overall ED for use with automated methods of estimating ingested volume.

The relative volume proportion of each food in an image was expressed as a number between 0 and 1 such that the sum of all food proportions for a given image always totaled to 1. For example, in a meal of chicken with peas and carrots, it was estimated that the chicken comprised about 1/3 of the total volume of the meal and, so, was entered as 0.33.
2.4. Food Exchange Method

This method involved analyzing food images based on the concept of food exchanges that are commonly used for meal planning by people with diabetes [29]. One choice can be exchanged for another in a specified amount within the same category because they are equivalent in terms of energy density and macronutrient composition. For example, 1 Starch choice = 1 Bread slice = 1/2 of a large ear of corn = 1/3 cup of cooked pasta. Based on standard exchange lists, one carbohydrate choice was defined as 15 grams of carbohydrate, one protein choice as 7 grams of protein, and one fat choice as 5 grams of fat [29]. An exchange reference list was developed to provide the number of choices of carbohydrate, fat, and protein per serving of common foods (S1) and included W for each food.

For each image analyzed, the operator allocated the relative volume proportion of each food item, then entered the number of carbohydrate, protein, and fat choices along with the W per food item using the exchange reference list. For example, 1 cup of 2% milk was listed as 1 protein choice, 1 fat choice, 1 CHO choice, and 0.40 water content.

\[
ED_{\text{exchange/J/g}} = \left[ \frac{(\text{CHO choice} \times 1.67) + (\text{PRO Choice} \times 1.67) + (\text{Fat Choice} \times 3.7)}{\text{Sum of Total Meal Choices}} \right] \times (1-W) \tag{3}
\]

where ED is energy density; CHO is carbohydrate CHO; PRO is protein; W is water content; and Atwater conversion of 1.67 is J/g for carbohydrate and protein and 3.77 is J/g for fat.

2.5. Food Score-Long and Food Score-Short Methods

The Food Score (FS) Method involved assigning fat, carbohydrate, and protein scores for every food in an image reflecting the relative contribution of each macronutrient towards the overall energy content of meal in the image. A FS reference list (S2, S3) was developed covering common foods which included W and assigned a macronutrient score, on a scale of 1 - 10, such that the score for each food totaled 10. For example, cooked rice was scored as 0 fat, 1 protein, 9 CHO, and 0.70 water content. The FS reference list was developed in both long and short versions (S2 and S3, respectively). The long version contained a comprehensive and exhaustive list of individual foods whereas the short version grouped foods of similar macronutrient composition (±2 g, 1 g, and 1 g per serving for carbohydrate, fat, and protein, respectively) in a condensed list.

For each image analyzed, the nutritionist assigned the relative volume proportion of each food item. Using the food score reference list (S2 or S3), the operator then entered only the fat score of each food item along with water score. Since the ED of carbohydrate and protein are equivalent, these scores represented the nonfat component and were calculated as 10 minus the fat score.

\[
ED_{\text{FS/J/g}} = \left[ \frac{(\text{Fat score} \times \text{proportion} \times 3.77) + ((10 - \text{fat score}) \times 1.67)}{10} \right] \times (1-W) \tag{4}
\]

where W is water content, Atwater conversion of 3.77 is J/g for fat, and 1.67 is...
2.6. Meal Method

For each food image analyzed, the nutritionist entered an estimated fat and \( W \) score for the meal, using the meal reference list (S4). As described in the FS method, the nonfat score accounted for remainder out of the total score of 10 and was derived by calculation. This method required no estimation of food proportions or serving sizes since the meal was analyzed as a whole.

\[
ED_{\text{meal}(t/g)} = \left[ \frac{(\text{Fat score} \times 3.77) + \left(\frac{(10 - \text{fat score}) \times 1.67}{10}\right)}{1-W} \right]
\]

where \( W \) is water content, Atwater conversion of 3.77 is J/g for fat, and 1.67 is J/g for carbohydrate and protein.

2.7. Comparison of Experimental ED Estimation Methods

This study was conducted in two separate phases: phase 1 was a feasibility test using a small number of images from a previous study \[19\] and four different experimental methods (Exchange, Food Score-Long, Food Score-Short, and Meal) for estimating ED to identify the optimal method (least time consuming and most accurate) whereas phase 2 was a full-scale validation of the optimal method identified in phase 1 compared with the WFR from a large database of dietary intake images. In phase 1, three trained nutritionists analyzed 116 food images that staff, not involved in this study, took of their own meals and uploaded in de-identified form to a secure server. Images, representative of all meals and snacks during the day in free living conditions, were randomly assigned to 4 sets (29 images/set) such that each had equal representation of breakfast, lunch, dinner, and snack images. One set of images was designated to each of the four methods: exchange, FSS, FSL, or meal method. Each of the three nutritionists analyzed images using all four experimental methods. The nutritionists were currently practicing in the field and were provided with training instructions for each method prior to analysis. An independent trained nutritionist, not involved in this study, coded the photographs, grouped them into representative sets of 29 images, and performed all FVM estimations and NDS-R entry.

In Phase 2, three trained nutritionists analyzed 213 images, derived from photographic food records collected as a part of a previous study \[19\], using only the meal method. An independent nutritionist analyzed the same photographs and conducted NDS-R entry using the FVM. The WFR were weighed and recorded by an independent nutritionist as part of the original study.

For all experimental methods, nutritionists were blind to the ED output to prevent them from changing data entry based on their perception of whether the estimated ED was correct. Nutritionists also entered the time(s) it took to analyze each food image.
2.8. Statistical Analysis

For phase 1, accuracy of the mean of three nutritionist estimates of energy density against the FVM was statistically analyzed using limit of agreement as discussed by Bland-Altman [30] (Table 1). Separate analysis was conducted for each of the four methods. Normality tests indicate log transformation to a natural base normalize the length of time. This transformed variable was compared across four methods using analysis of covariance (ANCOVA) adjusting for FVM time. Linear mixed effects model was used to perform this ANCOVA to account for correlation of repeated measures on the same sample. For phase 2, Bland-Altman analysis was also used to assess accuracy. To assess the overall accuracy among all nutritionists, a linear mixed effects model with compound symmetry covariance was used to model the difference between FVM, WFR and measure by each individual nutritionist and then estimate the SD of difference and bias. Inter-operator reliability among the three nutritionists was assessed using intra-class correlation coefficient. Individual reduction of time as compared to FVM were compared across the three nutritionists using non-parametric analysis of variance.

This study involves secondary data analysis. The parent study was powered on the precision of correlation for a sample size of 30. 30 participants were recruited with 28 completing the study. However, the size of precision (error margin) is about the same if the sample size is 28 vs 30 (0.647 for 38 vs 0.654 for 30) keeping confidence level (0.95) and correlation coefficient, 0.80, (95% C: 0.59 - 0.77) constant. Bootstrap of 5000 samples was used to calculate the 95% CI for inter-operator reliability coefficient. Post-hoc power analysis of the parent study showed that group sample sizes of 28 achieve 17% power to detect a difference of 0.14000 between the null hypothesis that both group correlations are 0.58000 and the alternative hypothesis that the correlation in group 2 is 0.44000 using a one-sided z test (which uses Fisher’s z-transformation) with a significance level of 0.05000.

3. Results

Phase 1 showed that, among experimental methods, the meal method showed the least variability and took significantly less analysis time per meal when compared to the other three methods (104 s vs 117 s vs 116 s vs 68 s for the exchange, FSL, FSS and meal methods, respectively; p = 0.03, Table 1). The meal method significantly decreased overall analysis time relative to the FVM (~120 s ± 16.4 s, p < 0.0001).

In phase 2, the images analyzed covered a broad range of foods, with EDs ranging from 1.5 to 20.9 J/g. The meal method generally over-estimated ED by 1.56 ± 3.17 J/g (p < 0.0001) compared to the FVM and 1.67 ± 3.09 J/g (p < 0.0001) and compared to the WFR (Table 2 and Figure 1). The meal method demonstrated strong inter-operator reliability as indicated by strong Intra-Class Correlation Coefficient (ICC) of 0.80.
Figure 1. Bland-Altman plots of mean difference of methods and ED. (a) FVM and WFR (b) Meal and WFR. Energy density of meals is x-axis, and the difference between the scores is y-axis. Parallel lines represent limits of agreement.

The major advantage of the meal method relative to the FVM was that it reduced analysis time by 69% per image (−120 s ± 16.4 s, p < 0.0001; Figure 2).

4. Discussion

In phase 1, all four experimental method significantly decreased overall analysis time relative to the FVM (Table 1). The meal method, however, was significantly faster when compared to the other three methods (p < 0.0001), and showed the least variability when compared to the full visual method. One main distinction between the meal method and other experimental methods is that the former did not require the operator to estimate relative proportions of each food within the meal. Consequently, the operator only had to focus on estimating fat and water content scores whereas other methods required estimation of multiple ingredients, portion size, water content, and macronutrient content. An increased number of variables included in a method increased the number of reference materials nutritionists had to review, thus increasing burden and time (Table 1). The lower number of variables in the meal method significantly decreased the analysis time per image (Table 1, p-value < 0.0001) and showed the lowest difference from the FVM estimate of ED.

In Phase 2, the FVM, as previously published [2], proved accurate relative to the WFR (difference between methods = 0.12 kJ/g ± 2.84, p = 0.54; Figure 1).
The faster meal method was less accurate than full visual estimation so further methodological improvements or mathematical correction are necessary when using this method. The meal method overestimated ED when compared to the visual method, 1.56 ± 3.17 J/g (p < 0.0001), and the WFR 1.67 ± 3.09 J/g (p < 0.0001; Table 2). This could potentially overestimate daily energy intake by about 555 kJ/d, or 6.8% based on the average daily energy intake of an American adult (8167 kJ; [27]). This contrasts strongly with other studies where food intake is generally underreported by at least 20% using standard self-report methods [1] [2] [3] [4] [5] [19] [31]. This is a distinct difference between the meal method and the FVM since inaccurate estimation of energy from fat has the greatest potential to skew dietary intake data in both adults and children and the meal method relies solely on accurate estimation of the fat content of a meal [19] [21].

The meal method showed no consistent pattern of food images where ED was inaccurately estimated. Some images in which ED was overestimated were single food items, such as an Oreo cookie, whereas other images contained several food items, such as a mixed meal of spaghetti with meat sauce and Sprite. This can be attributed to either poor knowledge of the fat content of certain foods by the nutritionists, inability to determine the exact food type from images (eg. skim milk vs full fat milk), or the fact that the meal reference list did not account for beverages combined with food items which made it difficult for nutritionists to estimate the water and fat content of the whole meal. This short-coming of the meal reference list will be corrected in future studies. Updating the meal reference list, including more detailed operator instructions, and standardized training sessions will increase inter-operator agreement by providing all nutritionists with proper knowledge on how to accurately and efficiently use the meal method.

Analysis time for the meal method was faster in Phase 2 (37 s ± 12) than in
Phase 1 (68s ± 25) which likely indicates a training effect as well as using a single method versus the four different methods in phase 1. The meal method reduced analysis time by 69% (120 s) per image relative to the FVM (Figure 2). This reduction in analysis time may seem small but could accumulate quickly. When comparing time spent on the FVM vs the meal method in this study, the total time saved for analyzing 217 food images would be 434 min per operator. Using the CCTSI Nutrition Core staff rate of $1.47/min, this represents a saving of $427.27 per operator.

The strengths of the present study include use of a WFR as the reference, use of trained, independent nutritionists for applying the FVM versus experimental methods, and a systematic study design. The strong ICC indicates that all three nutritionists implemented the meal method uniformly and with good fidelity to the protocol. Limitations include the limited food images and small number of total participants which decreased overall food variety of the images analyzed. Future studies should include a broader age range of participants and update the meal reference list and training instructions. It should be noted that the meal method can only be used to estimate ED and, therefore total energy intake, whereas the FVM can estimate energy intake as well as macronutrient and micronutrient content of the diet [1].

5. Conclusion

In conclusion, the meal method is a novel approach that can be used for analyzing food images to estimate ED and, thus, total energy intake from photographic food records and significantly decreases analysis time and cost compared to the FVM. Therefore, using the meal method could significantly decrease the cost of dietary intake measurements from food images, positively contributing towards affordability of the device use.

Acknowledgements

Contents are the authors’ sole responsibility and do not necessarily represent official NIH views. The authors wish to thank Kristen Bing, MS, RD, Nathalie Matamoros, and Stephanie Jung for data entry and analysis. We also acknowledge the support of the Colorado Clinical Translational Sciences Institute (CCTSI) Nutrition Core.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References


Supplementary Material

S3 File. Food Score Short Reference Sheet.