The test-retest reliability and validity of food photography and food diary analyses.

CURTIS, Christopher <http://orcid.org/0000-0002-4386-0284>, HILLS, Samuel P, ARJOMANDKHAH, Nicola, COOKE, Carlton, RANCHORDAS, Mayur <http://orcid.org/0000-0001-7995-9115> and RUSSELL, Mark

Section: Original Research

Article Title: The test-retest reliability and validity of food photography and food diary analyses

Authors: Christopher Curtis PhD, RSEN [1,2], Samuel P. Hills PhD [3], Nicola Arjomandkhah PhD [1], Carlton Cooke PhD [4], Mayur K. Ranchordas PhD, RSEN [5], Mark Russell PhD, ANutr [1]

Affiliations: [1] School of Sport and Wellbeing, Leeds Trinity University, United Kingdom; [2] School of Pharmacy & Nutrition, University of Navarra, Pamplona, Spain; [3] School of Social and Health Sciences, Bournemouth University, United Kingdom; [4] Leeds Beckett University, Headingley Campus, Leeds, United Kingdom; [5] Academy of Sport & Physical Activity, Health Research Institute and Advanced Wellbeing Research Centre, Sheffield Hallam University, United Kingdom

Corresponding Author: Christopher Curtis, School of Pharmacy & Nutrition, University of Navarra, Pamplona, Spain. Email: ccurtis@unav.es

Running Head: Reliability of dietary intake methods

Journal: Nutrition and Dietetics – Journal of Dieticians Australia

Abstract: 244

Word Count: 3352

Ethics Statement: The study obtained institutional ethical approval (Reference: SSHS-2017-082) and informed consent was sought from participants prior to study involvement.

Author Contributions: Conceptualization: CCu & MRu; Methodology: CCu, SH & MRu; Formal analysis: CCu; Investigation: CCu & SH; Resources: CCu & SH; Writing—original draft preparation: CCu & MRu; Writing—review and editing, CCu, SH, NA, CCo, MRa & MRu; Supervision: NA, CCo & MRu; Project administration: CCu & MRu. All authors have read and agreed to the published version of the manuscript.

Competing interests: The authors have no conflicts of interest, financial or otherwise, to declare.

Funding details: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

ABSTRACT:

Aims: To assess test-retest reliability of both food photography and food diary methods and validity of these data against known values derived from food labels.

Methods: Test-retest reliability analyses of food diary and food photography were compared using single foodstuffs using intra-class correlation coefficients, coefficients of variation and limits of agreement. For food diaries, 24-h test-retest reliability was also examined. Validity was assessed against weighed analyses. As part of habitual intake, a single foodstuff (randomly allocated from 14 common foods) were consumed by 26 participants over 24-h. On two occasions (14 days apart), single-blind dietary analyses allowed estimation of foodstuff-specific energy and macronutrient content, and 24-h intakes.

Results: For food diaries, test-retest reliability was acceptable (weight, energy, carbohydrate, protein, fat: all intraclass correlation coefficients >0.990, coefficient of variation percentage: <0.1%, limits of agreements: <0.1 to <0.1, p>0.05, effect size: <0.01). For food photography, test-retest reliability was acceptable for weight, energy, carbohydrate, and protein (all intraclass correlation coefficients >0.898, coefficient of variation percentage: 3.6% - 6.2%, limits of agreements: 1.1 to – 44.9, effect size: 0.01 – 0.12). Food photography validity was worse than food diaries for all variables (percentage difference: 8.8% - 15.3%, coefficient of variation percentage: 7.5% - 13.8%, all; p≤0.05, effect size: 0.001 – 0.11).

Conclusions: Greater reliability and validity occurred in food diaries versus food photography; findings which may suggest that using food photography may lead to an under-estimation of energy and macronutrient content, which may have implications for dietary interventions and nutritional strategies.

Keywords: Energy Intake; Energy Expenditure; Portion Size; Nutrition; Technology

Introduction:

Self-reported food diaries (FD) are frequently used to estimate food intake in a number of nutrition, clinical and sport settings, with participants reporting the portion sizes of their consumption with descriptions of the items [1]. Relative to the proposed optimal method of assessing dietary energy intake via doubly-labelled water methods and assumed balance between energy intake and total daily energy expenditure [1], methods based on FD are appealing from a cost and behaviour change perspective [1]. However, irrespective of how quantification of intake occurs in FD (i.e., household or weighed measures), participant burden and compliance, is still a consideration, particularly over longer periods of time (i.e., >7 days). Accordingly, this may cause under-reporting or selection bias as reported previously when FD have been implemented [2,3]. Food photography (FP) via smartphone nutrition application (app) technology has been proposed as an additional method of energy intake assessment [4,5], yet the validity and reliability of FP has not been comprehensively evaluated to date; research that will likely inform the decision-making of practitioners and researchers regarding the implementation of this method of energy intake estimation.

Early studies using nutrition app technology as a method of recording dietary intake focused primarily on clinical dietetic practice in specific populations (i.e., pregnant females, obese and paediatric individuals) [6,7,8], research of these technologies is emerging within healthy populations [9,10]. Although FP has been reported to provide valid estimates of energy intake [6], methods reliant solely on images can be prone to underestimation [11], with research suggesting a ~7-10% underestimation of total energy expenditure in free-living conditions [1,2,12]. Despite poor inter-practitioner reliability for FP analysis [12], it has been suggested that FP may enhance self-reporting by revealing unreported foods and allowing minimization of misreporting errors that are not captured by FD methods alone [1,2]. Within a clinical setting, several studies have investigated the validity of FP methodologies against weighed FD and 24-h dietary recall and doubly-labelled water techniques [8,9,13,14]. Despite showing acceptable limits of agreement (LoA) for various food groups, further investigation is required to compare the validity and reliability of FP and FD methods of determining energy and macronutrient content [14].

Use of FP technologies continues to grow within practice amongst nutrition practitioners with ~32% of sports nutrition practitioners globally using the method [15]. That said, research into the usage and engagement within this cohort warrants further investigation [16]. Simpson *et al.* [17] investigated preference to using such technologies, with participants preferring FP to FD due to perceived low burden [17]. Moderate increases in nutrition knowledge (~6%) were observed throughout an intervention that aimed to assess the feasibility of dietary education via mobile phone app use, together with any subsequent changes in nutrition and sports nutrition knowledge [17]. When comparing an estimated FD against FP (i.e., ‘Snap-n-Send’; [18]), Costello *et al.* [18] reported a small mean bias for under-reporting in FP, supporting the use of FP to assess diet accurately. Despite further commentary highlighting FP as a valid and reliable method of assessing EI within adolescent athletes [19], such methods require an examination of their reliability and validity (including energy and macronutrient content) against a criterion before these aspects of scientific investigation can be confirmed [20,21].

Given the limited literature in general populations using FP technologies, and the need to further assess the reliability and validity of FP and FD against weighed food intake as a method of assessing dietary and macronutrient intake, furthering such research will help to develop further understanding of energy balance (intake vs expenditure) and allow for practitioners to intervene, with a view to accurately recommending nutritional interventions in a range of cohorts, and provide improved accuracy in dietary intake monitoring. The aim of this study was therefore, to assess the test-retest reliability of both FP and FD methods and to assess the validity of these data against the known values derived from food labels.

Methods:

A total of 26 (9 males, 17 females) participants, recruited via poster advertisement, volunteered for the study and were free from allergies or intolerances to substances contained within the food items provided. Post-hoc analyses using the means of energy (megajoules; MJ) from the FP_{Item} versus FD_{Item}, with beta set at 0.80 and α being equal to 0.05 (two-tailed), yielded a statistical power of 0.97.. The study obtained institutional ethical approval (Reference: SSHS-2017-082) and informed consent was sought from participants prior to study involvement. This study was carried out in accordance with the Guidelines for Reporting Reliability and Agreement Studies (GRRAS) guidelines [22].

The test-retest reliability of the FD and FP analysis methods were compared using single foodstuffs (FD_{Item}, FP_{Item}) and in the case of FD, over a 24-h period also (FD_{24}). Validity was assessed using food label-informed weighed portion analyses. As part of habitual intake, a single foodstuff were consumed (randomly allocated from a list of 14 common foods based upon previous work from the research team; Table 3) on an individual basis by 26 participants over 24-h with requirements to record the intake via FD (i.e., self-reported recording) and FP (i.e., standardized photograph). All food consumed over the 24-h was also recorded by FD. On two occasions (14 days apart), single blind dietary software analyses allowed energy and macronutrient content estimation of single foodstuffs (FD_{Item} and FP_{Item}), and 24-h intake (FD_{24}).

Participants were invited to the laboratory for the purposes of familiarization of methods and to collect foodstuffs and items necessary to conduct the study. Participants were asked to consume the allocated foodstuff as a single meal/snack within their 24-h habitual dietary intake and record its consumption via the FD and FP methods outlined below. To ensure blinding, the lead researcher was unaware of the specific food item that was provided to participants as other members of the research team delivered this aspect of the study. Where required, participants were instructed on any storage and/or cooking/re-heating methods and were encouraged to consume the whole portion(s) provided to ensure accuracy and reduction in measurement error.

Following familiarization of methods (including a sample presentation of FD completion and FP instructions), participants were instructed to record their 24-h food and fluid intake via FD that incorporated quantity estimation via household measures (as per the methods of Russell & Pennock; [23]). Briefly, participants were required to provide detailed descriptions at the time of consumption (including timing of consumption, food description - referencing relevant cooking methods and/or brand names, estimated quantity prepared, estimated quantity remaining - if relevant, allowing calculation of consumed quantity, volume in the case of fluid consumed) of foodstuffs consumed in a 24-h period representing habitual practices from midnight to midnight between consecutive days. To provide standardized images against which the FP data could be compared, a variety of prepared foodstuffs were photographed by a member of the research team, on an individual basis using a digital camera (Nikon D3500, Nikon, Tokyo, Japan) in natural light on a white dining plate (plate dimension = 26 cm, height of photograph = 50 cm, angle of photograph = 45 degrees; [20]) and place mat (A3: 42 cm x 30cm). For drinks, a standard height glass (capacity: 340ml; dimension: 15cm x 6.4cm) was used.

Participants were familiarized with the FP standardization method and were instructed to take a photo of the provided foodstuff according to standardized instructions (i.e., photograph height: 50 cm, angle of photograph: 45º, use of a provided calibrated place mat: A3; 42 x 30 cm with grid squares measuring 7cm x 6 cm for size standardization) using a camera or mobile phone or tablet device with a camera function. Where possible, participants were requested to use a white plate, take photos in natural light, and position the plate in the centre of the place mat provided. In an event of a portion of food not being fully consumed, participants were asked to photograph any remaining content according to the preceding instructions. All images captured were required to be sent to the lead author via email within 24-h of images being taken. Once all analysis methods were completed, the lead researcher was unblinded to the identity of the foodstuffs provided to each participant.

Using energy and macronutrient content derived from dietary software analyses (Nutritics v.4.3.15), the test-retest reliability of FD_{Item}, FD_{24} and FP_{Item} was investigated. Analyses were conducted by the lead author and stored on a single spreadsheet that was sent to other members of the research team post-analysis for a period of two weeks. During this time, no data was kept by the lead researcher to minimize recollection bias between repeated measurements of dietary intake [24]. After two weeks, the original FD and FP sources were re-analysed by the lead researcher to determine the test (T_1) and retest (T_2) reliability of each method. These analyses were then checked by the principal investigator.

Following preparation according to their cooking guidelines, 14 commonly consumed foods and beverages were assessed by a member of the research team (to ensure blinding of the lead author) in terms of label-informed weighed portion analyses using digital food scales (Salter Disc Electronic Scales). The data generated from weighed portion analyses was the comparator information which was deemed the "standard" against which subsequent FP and FD analyses were compared.

Data were analysed via SPSS (SPSS Statistics for Windows, Version 25.0, Armonk, NY: IBM Corp). Normality was assessed via Shapiro-Wilks test. For reliability analyses, paired t-tests assessed between-trial differences for energy and macronutrient content (both FD and FP) and estimation of foodstuff weight from both FD and FP. Intra-class correlation coefficients (ICC: two-way mixed method, absolute agreement) were used to assess the test-retest reliability of repeated energy and macronutrient content derived from FP and FD. Coefficients of variation (CV%; typical error expressed as a percentage of the subject's mean score), and LoA (mean bias±1.96 standard deviation; SD) were calculated and provided that no significant differences existed, variables were deemed to have acceptable reliability if both CV% was ≤10% and ICC was ≥0.8 [25]. Bland-Altman scatterplots were produced to evaluate potential bias between mean differences [26]. For validity, the FD_{Item} and FP_{Item} analyses were compared against a weighed portion analyses criterion by mean difference percentage (%diff), coefficient of variation percentage (CV%), LoA, p-values and effect sizes (ES). Effect sizes were calculated in accordance with Cohen's *d* ES principles (0<ES<0.2: Trivial, 0.2<ES<0.5: Small, 0.5<ES<0.8: Medium, >0.8: Large; [27]). An alpha level of p≤0.05 denoted significance.

Results:

Acceptable test-retest reliability was observed for all variables for FD_{24} (energy: ICC: 0.999, CV%: 0.4±0.7, LoA: 66.1 to -52.2, ES: 0.01; carbohydrate: ICC: 0.999, CV%: 0.5±0.9, LoA: 9.4 to -8.8, ES: 0.01; protein: ICC: 0.998, CV%:0.2±0.5, LoA: 5.5 to -3.7, ES: 0.001; fat: ICC: 0.999, CV%: 0.1±0.3, LoA: 2.1 to -1.5, ES: 0.001, all $p \geq 0.05$; Table 1; Figure 1). Visual representation of the Bland-Altman scatterplots suggests that for energy, carbohydrate, protein and fat within FD_{24}, thresholds of mean bias and LoA's are absent (Figure 1). Acceptable test-retest reliability was also observed for all variables derived from the FD_{Item} analyses (weight, energy, carbohydrate, protein, fat: all ICC: >0.99, CV% <0.1%, LoA: <0.1 to <0.1, ES: 0.01, all $p \geq 0.05$; Table 2). However, while acceptable test-retest reliability was observed in FP_{Item} for weight (ICC: 0.998, CV%: 6.2±9.4, LoA: 26.8 to -25.3, $p>0.05$, ES: 0.01), energy (ICC: 0.973, CV%: 5.1±8.6, LoA: 42.8 to -44.9, $p>0.05$, ES: 0.02), carbohydrate (ICC: 0.898, CV%: 4.7±6.6, LoA: 16.8 to -20.4, $p>0.05$, ES: 0.12), and protein (ICC: 0.998, CV%: 3.8±5.3, LoA: 1.4 to -1.6, ES: 0.01, $p>0.05$; Table 2), this was not the case for fat (ICC=0.993, CV%: 11.7±30.8, LoA: 0.7 to -0.9, ES: 0.03, $p \geq 0.05$; Table 2; Figure 2). Visual representation of the Bland-Altman scatterplots suggests that for weight, energy, carbohydrate, protein and fat within the FP_{Item}, thresholds of mean bias and LoA's are absent (Figure 2).

INSERT TABLE 1 ABOUT HERE

INSERT FIGURE 1 ABOUT HERE

INSERT TABLE 2 ABOUT HERE

INSERT FIGURE 2 ABOUT HERE

On a foodstuff-specific basis (Table 3), FP_{Item} demonstrated worse levels of agreement than FD_{Item} for all variables (Table 4).

INSERT TABLE 3 ABOUT HERE

INSERT TABLE 4 ABOUT HERE

Discussion:

The primary aim of this study was to assess the test-retest reliability of the FD and FP methods for analysing single foodstuffs, and in the case of FD, dietary intake collected over 24-h. A further aim was to assess the validity of each method against label-derived weighed portion analyses for single foodstuffs. Analyses of single foodstuffs via the FD and FP methods showed that FD_{Item} demonstrated acceptable test-retest reliability for all variables, whereas FP_{Item} did not show acceptable reliability for estimating fat intake. Likewise, FD_{24} showed acceptable reliability for all energy and macronutrient content variables, this statement is further substantiated by the absence of thresholds mean bias and respective LoA's (Table 1 and Figure 1). Regarding validity, all variables assessed via FD had stronger agreement than those assessed by FP, with agreement between weighed portion analyses and FD being strongest for carbohydrate relative to all other variables. When considering the analysis methods alone, relative to FD, using FP may lead to an under-estimation of energy and macronutrient content, which may have implications for subsequent conclusions and/or nutritional strategies that may be informed by data provided during dietary intake and subsequent analysis via FD or FP. Therefore, this study offers practitioners and researchers an insight into the reliability and validity of the FD and FP methods, suggesting that foodstuff-specific analyses favouring FD.

The absence of statistically statistical differences and trivial ES between repeated estimations of energy and macronutrient content, combined with the threshold of reliability used here, demonstrated acceptable levels of test-retest reliability for FD_{24} (Figure 1); a finding which supports the data of Ortega *et al*. [28] who reported that FD_{24}

elicited reliable and precise estimates of energy intake [28]. Additionally, when considered in terms of ICC, our findings indicated greater reliability in FD_{24} for EI compared to those observed by Putz *et al*. [29]. Likewise, the mean CV values for energy (CV% Energy Intake $MJ \cdot d^{-1}$: 0.4±0.7) and macronutrient (CV% CHO $g \cdot d^{-1}$: 0.5±0.9, CV% PRO $g \cdot d^{-1}$: 0.2±0.5, CV% FAT $g \cdot d^{-1}$: 0.1±0.3) content within FD_{24} methods observed in this study showed greater reliability than those observed by Braakhius *et al*. [30], being 33% - 57%, when the same dietician undertook a test-retest design of 24-h FD analyses. While the reason for such discrepancies remains unclear, it is possible that the use of the analysis tool influences the data achieved as studies have previously used pre-coded, paper-based food records [29] or differing databases [30] to that presented here.

The energy and macronutrient content findings in the present study showed comparable reliability within FP methods to those observed by Martin *et al.* [1]. No statistical differences were observed across all variables in the test-retest reliability analyses for FP in the present study, which are further supported by the observed marginal ES (Table 2). While such findings may be influenced by several factors, advances in mobile phone camera technology may have improved picture quality and subsequent precision of analyses. Research indicates that mobile phone devices with a ≥5 megapixel (MP) camera are capable of greater resolution imaging over differing magnifications [31] compared to the 1.3 MP mobile device adopted by Martin *et al.* [1]. Mobile phone devices such as the iPhone 4 (released in 2010) contain a minimum of a 5 megapixel camera [31], with all participants using a similar (or more modern) device.

When compared to weighed portion analyses, our validity findings indicate that FD_{Item} offers better levels of agreement than FP_{Item}. Energy content derived from FD offered stronger levels of agreement with weighed portion analyses compared to FP analyses. Our findings offer comparable results to that of Fuller *et al*. [31] who compared agreement between a written and electronic FD (FDE). Despite observing similar mean differences between the two methods, the authors reported wide LoA between the two methods of FD for mean energy intake (FD: 7541.0±1966.0 vs. FDE: 7452.0±2024.0, LoA: −1848.0 to 2019.0), carbohydrate (FD: 186.0±58.0 vs. FDE: 178.0±60.0, LoA: −46.0 to 62.0), protein (FD: 91.0±27.0 vs. FDE: 94.0±28.0, LoA: −44.0 to 38.0) and fat (FD: 67.0±22.0 vs. FDE: 67.0±23.0, LoA: −30.0 to 30.0). Moreover, Costello *et al*. [18] compared an estimated FD against a FP method (i.e., 'Snap-n-Send') and reported a small mean bias for under-reporting across 96-h (−0.75 $MJ \cdot day^{-1}$; −5.7% to −2.2%, p<.001), 72 h (−0.76 $MJ \cdot day^{-1}$; −5.6% to −2.1%, p=.004) and 10-h (−0.72 $MJ \cdot day^{-1}$; −8.1% to −0.1%; p=.067). These findings, coupled with those in the present study, indicate that when using FP to assess energy intake, underestimation occurs irrespective of timeframe of sample from single items to 96-h intakes.

Despite both methods under-estimating energy and macronutrient content when compared with weighed portion analyses across both methods (FD_{Item}: ~11% - 18%; FP_{Item}: ~16% - 24%; Table 4), foodstuff specific comparisons were in better agreement with weighed portion analyses when assessed via FD versus FP. Small ES were observed across all comparisons to weighed portion analyses irrespective of method (FD_{Item} and FP_{Item}) and notwithstanding the lack of significance, larger percentage differences and CV values existed for FP_{Item} versus FD_{Item}. These findings infer that FP methodologies, with the absence of thresholds for mean bias and LoA's, demonstrate less agreement (Table 2 and Figure 2) when assessing energy and macronutrient content compared to FD. From an applied perspective, the findings of the current study offer insights into the validity and reliability of methods to assess energy and macronutrient content. Despite both methodologies underestimating values, the magnitude of differences from FD_{Item} compared to FP_{Item} are lower. Our findings highlighted larger discrepancies in values between methods on 'loose' items (e.g., Baby Potatoes, Microwavable Rice, Breakfast Cereals etc.; Table 3) from FP when compared to FD against the weighed analyses comparator. Such findings suggest that caution must be exercised when loose items are the basis of investigation [33] as under-estimation of such items, both individually and as part of a meal, have been demonstrated during analysis undertaken by experienced (i.e., >3 years' experience) and inexperienced (i.e., recent graduates) nutrition practitioners [12].

This study, whilst providing novel data on the test-retest reliability and validity of both FP and FD methods when assessing the energy and macronutrient content of single foodstuffs in adult populations, is not without limitations. Firstly, despite our findings indicating varying levels of agreement for the FP and FD methods regarding total energy and macronutrient analysis, micronutrient assessment was not undertaken and thus remains to be considered from a reliability and validity perspective. Additionally, an extended FD and FP data collection period (e.g., 3 d, 7 d) may have provided greater insight into the reliability and validity of both FP and FD methods allowing greater comparison to previous research [1,18,28,30]. However, real-time recording methods (e.g., FD) may be a source of error for participants to record dietary intake accurately, in addition to limitations in dietary analysis software databases [1,29]. Similarly, with increased duration of timeframe, increases participant burden and/or risk of dropout increases [33]. To limit effects of such errors, the current study requested a 24-h period of dietary intake recording incorporating a randomly allocated foodstuff to incorporate the ability of a test-retest of FD and FP. Our post-hoc analyses from the means of energy (MJ) from the FP_{Item} versus FD_{Item} determined statistical power at 0.97 which was preserved with the timeframe adopted within the present study.

In conclusion, FD_{Item} showed greater test-retest reliability and validity compared with FP_{Item}, despite both methods under-estimating values by 11 – 24%. Except for fat intake derived from FP_{Item}, both FD_{Item} and FP_{Item} showed acceptable reliability across all variables when test-retest analyses were undertaken. Notably, a comparison of the two methods shows acceptable CV% values (CV%: 7.5 – 9.3; Table 4) and trivial ES (0.001 – 0.100) between methods for weight, energy, carbohydrate, and protein. However, while these findings indicate that both methods may be reliable at assessing these variables, validity relative to a weighed analyses comparator, was less in FP_{Item} versus FD_{Item}. Accordingly, these findings suggest that FD may be considered a reliable and valid tool for nutrition practitioners and researchers to assess dietary intake of adults in free-living conditions, when compared against FP. Therefore, it could be suggested that in situations where accurate energy intake and macronutrient analysis are required, then FD should be used as opposed to FP.

References:

1. Martin CK, Han H, Coulon SM, et al. A novel method to remotely measure food intake of free-living people in real-time: The Remote Food Photography Method (RFPM). *Br J Nutr*, 2009; 101:446–456. doi: 10.1017/S0007114508027438.
2. Martin CK., Nicklas T, Gunturk B, et al. Measuring food intake with digital photography. *J Hum Nutr Diet*, 2014; 27:72-81. doi: 10.1111/jhn.12014
3. Rosenbloom CA, Loucks AB, Ekblom B. Special populations: The female player and the youth player. *J. Sport Sci*. 2006; 24:783-793. doi: 10.1080/02640410500483071
4. Foster E, Hawkins A, Barton KL, et al. Development of food photographs for use with children aged 18 months to 16 years: Comparison against weighed food diaries – The Young Person's Food Atlas (UK). *PLoS ONE*, 2017; 12:e0169084. doi: 10.1371/journal.pone.0169084
5. Naska A, Valanou E, Peppa E, et al. Evaluation of a digital food photography atlas used as portion size measurement aid in dietary surveys in Greece. *Public Health Nutr*, 2016; 19:2369-2376. doi: 10.1017/S1368980016000227
6. Lieffers JR, Vance VA, Hanning, RM. Use of mobile device applications in Canadian dietetic practice. *Can. J. Diet. Pract*. Res, 2014; 75:41-47. doi: 10.3148/75.1.2014.41
7. Thomas JG, Wing RR. (2013) Health-e-call, a smartphone-assisted behavioural obesity treatment: pilot study. *JMIR Mhealth Uhealth*, 2013; 1:E3. doi: 10.2196/mhealth.2164

8. Ashman AM, Collins CE, Brown LJ, et al. Validation of a Smartphone Image-Based Dietary Assessment Method for Pregnant Women. *Nutrients*. 2017; 18;9(1):73. doi: 10.3390/nu9010073
9. Martin CK, Correa JB, Han H, et al. Validity of the Remote Food Photography Method (RFPM) for Estimating Energy and Nutrient Intake in Near Real-Time. *Obesity*. 2012; 20:891–899. doi: 10.1038/oby.2011.344
10. Lassen AD, Poulsen S, Ernst L, et al. Evaluation of a digital method to assess evening meal intake in a free-living adult population. *Food Nutr. Res*. 2010; 54:5311. doi: 10.3402/fnr.v54i0.5311
11. Gemming L. Utter J, Mhurchu CN. Image-assisted Dietary Assessment: A Systematic Review of the Evidence. *J Acad Nutr Diet*, 2015; 115:64-77. doi: 10.1016/j.jand.2014.09.015
12. Stables RG, Kasper AM, Sparks AS, Morton JP, Close GL. An Assessment of the Validity of the Remote Food Photography Method (Termed Snap-N-Send) in Experienced and Inexperienced Sport Nutritionists. *Int J Sport Nutr Exerc Metab*, 2021: 31:125-134. doi: 10.1123/ijsnem.2020-0216
13. Kong K, Zhang L, Huang L, et al. Validity and practicability of smartphone-based photographic food records for estimating energy and nutrient intake. *Asia Pac J Clin Nutr*, 2017; 26:396-401. doi: 10.6133/apjcn.042016.05
14. Bouchoucha M, Akrout M, Bellali H, et al. Development and validation of a food photography manual, as a tool for estimation of food portion size in epidemiological dietary surveys in Tunisia. *Libyan J Med*, 2016; 11:32676. doi: 10.3402/ljm.v11.32676

15. Jospe MR, Fairbairn KA, Green P, et al. (2015) Diet App Use by Sports Dieticians: A Survey in Five Countries. *JMIR Mhealth Uhealth*, 2015; 3:E7. doi: 10.2196/mhealth.3345
16. Curtis C, Russell M, Ranchordas MK (2018) Enhancing dietary practices, general nutrition knowledge and body composition of a female International Rugby Union player incorporating smartphone application technology. *J Sports Med and Phys Fit*, 2018; 58:366-368. doi: 10.23736/S0022-4707.17.07623-X
17. Simpson A, Gemming L, Baker D, et al. Do Image-Assisted Mobile Applications Improve Dietary Habits, Knowledge, and Behaviours in Elite Athletes? A Pilot Study. *Sports (Basel)*, 2017; 5:E60. doi: 10.3390/sports5030060.
18. Costello N, Deighton K, Dyson J, et al. Snap-N-Send: A valid and reliable method for assessing the energy intake of elite adolescent athletes. *Eur J Sport Sci*, 2017; 17:1044-1055. doi: 10.1080/17461391.2017.1337815
19. Costello N, McKenna J, Deighton K, et al. Commentary: Snap-N-Send: A Valid and Reliable Method for Assessing the Energy Intake of Elite Adolescent Athletes. *Front Nutr*, 2017; 4:47. doi: 10.3389/fnut.2017.00047
20. Rollo ME, Williams RL, Burrows T, Kirkpatrick SI, Bucher T, Collins CE. What are they really eating? A review on new approaches to dietary intake assessment and validation. *Curr Nutr Rep*, 2016; 5:307–314. doi:10.1007/s13668-016-0182-6
21. Kirkpatrick SI, Subar AF, Krebs-Smith SM, Thompson FE, Reedy J, Schap TE, et al. Evaluation of dietary assessment tools: does "validated" mean what we think it means? *FASEB J*, 2016; 30:43.8. doi: 10.1096/fasebj.30.1_supplement.43.8

22. Kottner J, Audigé L, Brorson S, Donner A, Gajeweski BJ, Hróbjartsson A, *et al*. Guidelines for reporting reliability and agreement studies (GRRAS) were proposed. *J Clin Epidemiol*, 2011; 64(1):96-106. doi: 10.1016/j.jclinepi.2010.03.002

23. Russell M., Pennock A. Dietary analysis of young professional soccer players for 1 week during the competitive season. *J Strength Cond Res*, 2011; 25:1816 – 1823. doi: 10.1519/JSC.0b013e3181e7fbdd

24. Marx RG, Menezesb A, Horovitza L, et al. A comparison of two time intervals for test-retest reliability of health status instruments. *J Clin Epidemiol*, 2003; 56:730-735. doi: 10.1016/s0895-4356(03)00084-2

25. Aben HGJ, Hills SP, Higgins D, et al. The Reliability of Neuromuscular and Perceptual Measures Used to Profile Recovery, and the Time-Course of Such Responses Following Academy Rugby League Match-Play. *Sports*, 2020; 8:73. doi: 10.3390/sports8050073

26. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*, 1986; 1:307–310.

27. Hopkins WG, Marshall SW, Batterham AM, et al. Progressive statistics for studies in sports medicine and exercise science. *Med Sci Sports Exerc*, 2009: 41:3–13. doi: 10.1249/MSS.0b013e31818cb278

28. Ortega RM, Pérez-Rodrigo C, López-Sobaler AM. Dietary Assessment Methods: Dietary Records. *Nutr Hosp*, 2015; 3:38-45. doi: 10.3305/nh.2015.31.sup3.8749

29. Putz P, Kogler B, Bersenkowitsch I. Reliability and validity of assessing energy and nutrient intake with the Vienna food record: a cross-over randomised study. *Nutr J*, 2019; 18:7. doi: 10.1186/s12937-019-0431-9

30. Braakhuis AJ, Stewart KL, Cox G, et al. Variability in Estimation of Self-reported Dietary Intake Data from Elite Athletes Resulting from Coding by Different Sports Dietitians. *Int J Sport Nutr Exerc Metab*, 2003; 13:152-165. doi: 10.1123/ijsnem.13.2.152

31. Skandarajah A, Reber CD, Switz NA, et al. Quantitative Imaging with a Mobile Phone Microscope. *PLoS ONE*, 2014; 9:e9690. doi: 10.1371/journal.pone.0096906

32. Fuller NR, Fong M, Gerofi J, et al. Comparison of an electronic versus traditional food diary for assessing dietary intake – A validation study. *Obes Res Clin Pract*, 2017; 11:647-654. doi: 10.1016/j.orcp.2017.04.001

33. Naska A, Lagiou A, Lagiou P. Dietary assessment methods in epidemiological research: current state of the art and future prospects. *F1000Res*, 2017; 6:926. doi: 10.12688/f1000research.10703.1

Table 1: Mean test (T_1) and re-test (T_2) reliability of dietary analysis for mean energy ($MJ \cdot d^{-1}$) and macronutrient intake ($g \cdot d^{-1}$) measured by 24-h food diary.

	T_1 (±SD)	T_2 (±SD)	Mean Diff (%)	ICC	Mean CV (%)	LoA	*ES*
Energy Intake ($MJ \cdot d^{-1}$)	9.13±2.83	9.16±2.79	0.4±1.1	0.999	0.4±0.7	66.1 to -52.2	0.010
CHO ($g \cdot d^{-1}$)	228±92	229±91	0.7±1.3	0.999	0.5±0.9	9.4 to -8.8	0.010
PRO ($g \cdot d^{-1}$)	113±39	113±40	0.7±1.6	0.998	0.2±0.5	5.5 to -3.7	0.001
FAT ($g \cdot d^{-1}$)	88±34	88±33	0.5±1.1	0.999	0.1±0.3	2.1 to -1.5	0.001

Mean Diff%, Mean Difference %; ICC, Intra-class correlation coefficients; Mean CV, Mean coefficient of variation; LoA, limits of agreement; ES, Effect Size; CHO, Carbohydrates; FAT, Fats; MJ, Megajoules; PRO, Proteins; SD, Standard deviation. Absence of symbols between T_1 and T_2 data points indicates no significant differences ($p>0.05$)

Table 2. Test (T_1) and re-test (T_2) reliability of single foodstuff mean weight (g), energy (MJ), and macronutrient intake (g) measured by food photography and food diary methods.

	FP_{Item}							*FD_{Item}*						
	T_1 (±SD)	T_2 (±SD)	Mean Diff (%)	ICC	Mean CV (%)	LoA	*ES*	T_1 (±SD)	T_2 (±SD)	Mean Diff (%)	ICC	Mean CV (%)	LoA	*ES*
Weight (g)	119.4±76.8	120.1±76.7	0.6±9.7	0.998	6.2±9.4	26.8 to -25.3	0.01	120.8±71.5	120.8±71.5	0.0	>0.99	0.0	0.0 to 0.0	0.01
Energy (MJ)	0.52±0.29	0.51±0.28	7.7±11.7	0.973	5.1±8.6	42.8 to -44.9	0.02	0.54±0.28	0.54±0.28	0.0	>0.99	0.0	0.0 to 0.0	0.01
CHO (g)	22.6±17.3	20.8±13.7	7.6±13.8	0.898	4.7±6.6	16.8 to -20.4	0.12	22.4±14.1	22.4±14.1	0.0	>0.99	0.0	0.0 to 0.0	0.01
PRO (g)	6.0±8.4	5.9±8.4	6.4±10.9	0.998	3.8±5.3	1.4 to -1.6	0.01	5.7±8.0	5.7±8.0	0.0	>0.99	0.0	0.0 to 0.0	0.01
FAT (g)	2.0±3.8	1.9±3.7	5.0±23.5	0.993	11.7±30.8	0.7 to -0.9	0.03	2.0±3.7	2.0±3.7	0.0	>0.99	0.0	0.0 to 0.0	0.01

Mean Diff%, Mean Difference %; ICC, Intra-class correlation coefficients; Mean CV, Mean coefficient of variation; LoA, limits of agreement; ES, Effect Size; CHO, Carbohydrates; FAT, Fats; FD, Food Diary; FP, Food Photography; MJ, Megajoules; PRO, Proteins; SD, Standard deviation. Absence of symbols between two data points indicates no significant differences ($p>0.05$)

Table 3. Absolute data of weighed analysis compared to mean ± standard deviation of estimated dietary analysis of foodstuffs from participant food photography and food diary methods**.**

Food Stuff	**WA**					**FP**					**FD**				
	Weight (g)	Energy (MJ)	CHO (g)	PRO (g)	FAT (g)	Weight (g)	Energy (MJ)	CHO (g)	PRO (g)	FAT (g)	Weight (g)	Energy (MJ)	CHO (g)	PRO (g)	FAT (g)
Banana	116.0	0.35	20.3	1.3	0.0	113.0±22.3	0.35±0.005	19.7±0.3	1.2±0.1	0.1±0.0	90.0±14.1	0.34±0.07	19.3±4.6	1.1±0.2	0.2±0.2
Crumpet	116.0	0.98	48.1	8.0	1.2	77.5±0.0	0.65±0.0	32.2±0.0	5.4±0.0	0.8±0.0	80.0±0.0	0.67±0.0	33.2±0.0	5.5±0.0	0.8±0.0
Bagel	81.0	0.90	42.9	8.1	1.5	90.0±0.0	1.00±0.0	48±0.0	9±0.0	1.6±0.0	79.0±4.2	0.91±0.002	43.0±0.0	8.2±0.0	1.5±0.1
Apple	132.0	0.25	13.1	0.7	0.5	105.0±0.0	0.23±0.0	12.1±0.0	0.63±0.0	0.53±0.0	105.0±0.0	0.23±0.0	12.2±0.0	0.6±0.0	0.5±0.0
Blueberries	74.0	0.12	6.7	0.7	0.2	38.0±13.4	0.06±0.02	3.5±1.2	0.3±0.1	0.1±0.0	75.0±35.4	0.13±0.06	6.9±3.2	0.7±0.3	0.2±0.1
Cereal bar	45.0	0.92	20.9	2.5	14.0	45.0±0.0	0.92±0.0	20.9±0.0	2.5±0.0	14±0.0	45.0±0.0	0.92±0.0	20.9±0.0	2.5±0.0	14.0±0.0
Soya Milk (chocolate flavour)	259.0	0.64	19.9	8.0	4.7	250.0±0.0	0.61±0.0	19.3±0.0	7.8±0.0	4.5±0.0	250.0±0.0	0.61±0.0	19.3±0.0	7.8±0.0	4.5±0.0
Apple Juice (cloudy)	264.0	0.42	25.3	0.3	0.0	250.0±0.0	0.40±0.0	25.0±0.0	0.3±0.0	0.0±0.0	250.0±0.0	0.40±0.0	25.0±0.0	0.3±0.0	0.0±0.0
Wholegrain Rice (Microwaveable)	247.0	1.65	75.3	9.1	6.4	108.8±12.4	0.66±0.07	40.7±16.9	4.9±1.4	2.1±0.3	125.0±63.6	0.84±0.53	38.9±19.9	4.7±2.3	3.2±1.6
Baby Potatoes	373.0	1.03	54.8	6.3	0.4	167.5±10.6	0.43±0.02	22.8±1.4	3.0±0.2	0.2±0.0	170.0±42.4	0.44±0.1	23.1±5.8	3.1±0.3	0.2±0.0
Italian-style Salad Leaves	35.0	0.02	0.7	0.5	0.2	31.5±2.1	0.02±0.001	0.6±0.0	0.5±0.0	0.1±0.0	42.5±10.6	0.03±0.007	0.9±0.2	0.6±0.2	0.2±0.0
Corn Flakes (fortified)	48.0	0.74	39.9	3.4	0.4	37.0±14.8	0.51±0.15	30.9±12.5	2.6±1.0	0.3±0.1	50.0±0.0	0.76±0.0	42.0±0.0	3.6±0.0	0.4±0.0
Baked Beans (canned)	189.0	0.63	26.3	9.5	1.0	205.0±0.0	0.68±0.0	28.5±0.0	10.3±0.0	1.0±0.0	205.0±0.0	0.68±0.0	28.5±0.0	10.3±0.0	1.0±0.0
Tuna (in brine) drained	105.0	0.47	0.0	26.1	1.1	130.0±0.0	0.58±0.0	0.0±0.0	32.4±0.0	1.3±0.0	125.0±35.4	0.56±0.15	0.0±0.0	31.2±8.8	1.3±0.4

CHO, Carbohydrates; FAT, Fats; FD, Food Diary; FP, Food Photography; MJ, Megajoules; PRO, Proteins; SD, Standard Deviation; WA, Weighed Analysis

Table 4. Validity of weight (g), energy (MJ) and macronutrient intake (g) between foodstuffs assessed by weighed analysis, food photography and food diary. Mean difference percentage, coefficient of variation percentage, limits of agreement and effect sizes (ES) are presented.

Variable	Method			WA vs. FP_{Item}				WA vs. FD_{Item}				FP_{Item} vs FD_{Item}			
	WA	FP	FD	% diff	CV%	LoA	*ES*	% diff	CV%	LoA	*ES*	% diff	CV%	LoA	*ES*
Weight (g)	148.8±102.0	117.8±75.1	120.8±71.5	18.6±20.1	16.3±20.0	93.7 to -155.9	0.36	14.6±17.6	12.4±17.0	92.0 to -148.1	0.33	11.1±14.7	9.3±13.3	30.2 to -24.1	0.04
Energy (MJ)	0.65±0.42	0.51±0.28	0.54±0.28	18.3±21.7	16.4±21.9	110.4 to -178.7	0.41	12.7±18.2	10.9±17.9	98.4 to -152.6	0.35	10.6±15.6	9.0±14.3	47.5 to -33.4	0.11
CHO (g)	28.1±21.6	21.7±14.2	22.3±14.0	15.6±20.1	13.9±19.6	18.0 to -30.9	0.36	11.0±18.5	9.9±18.1	19.2 to -30.7	0.33	8.8±14.7	7.5±13.4	7.5 to -6.2	0.04
PRO (g)	6.0±6.8	5.7±8.3	5.7±8.0	17.7±18.2	14.4±17.9	4.5 to -5.0	0.04	13.6±16.9	10.7±16.3	3.9 to -4.6	0.04	9.5±14.3	8.0±13.2	0.9 to -1.1	0.001
FAT (g)	2.2±3.9	1.9±3.7	1.9±3.7	23.9±31.6	25.9±40.7	1.9 to -2.6	0.08	18.0±29.7	19.8±39.1	1.5 to -2.0	0.08	15.3±20.5	13.8±19.9	0.7 to -0.5	0.001

MJ, Megajoules; % diff, percentage difference; CV%, coefficient of variation percentage; FD, Food Diary; FP, Food Photography; LoA, limits of agreement; ES, effect size; WA, Weighed Analysis. Absence of symbols between two data points indicates no significant differences

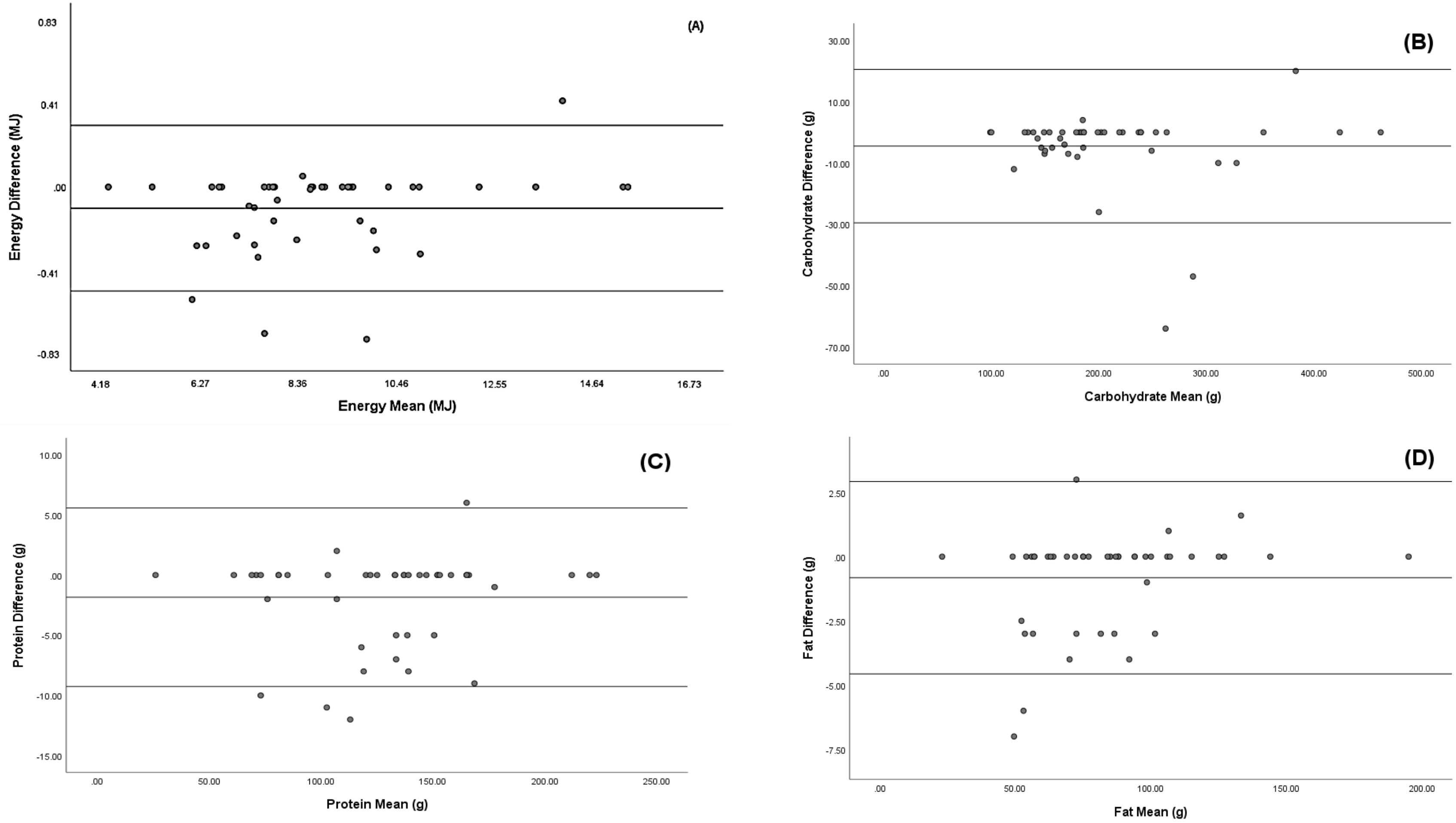

Figure 1: Bland-Altman scatterplots with limits of agreement (mean bias ± 1.96 SD) for test-retest reliability of 24-h food diaries for Energy (A), Carbohydrate (B), Protein (C) and Fat (D).

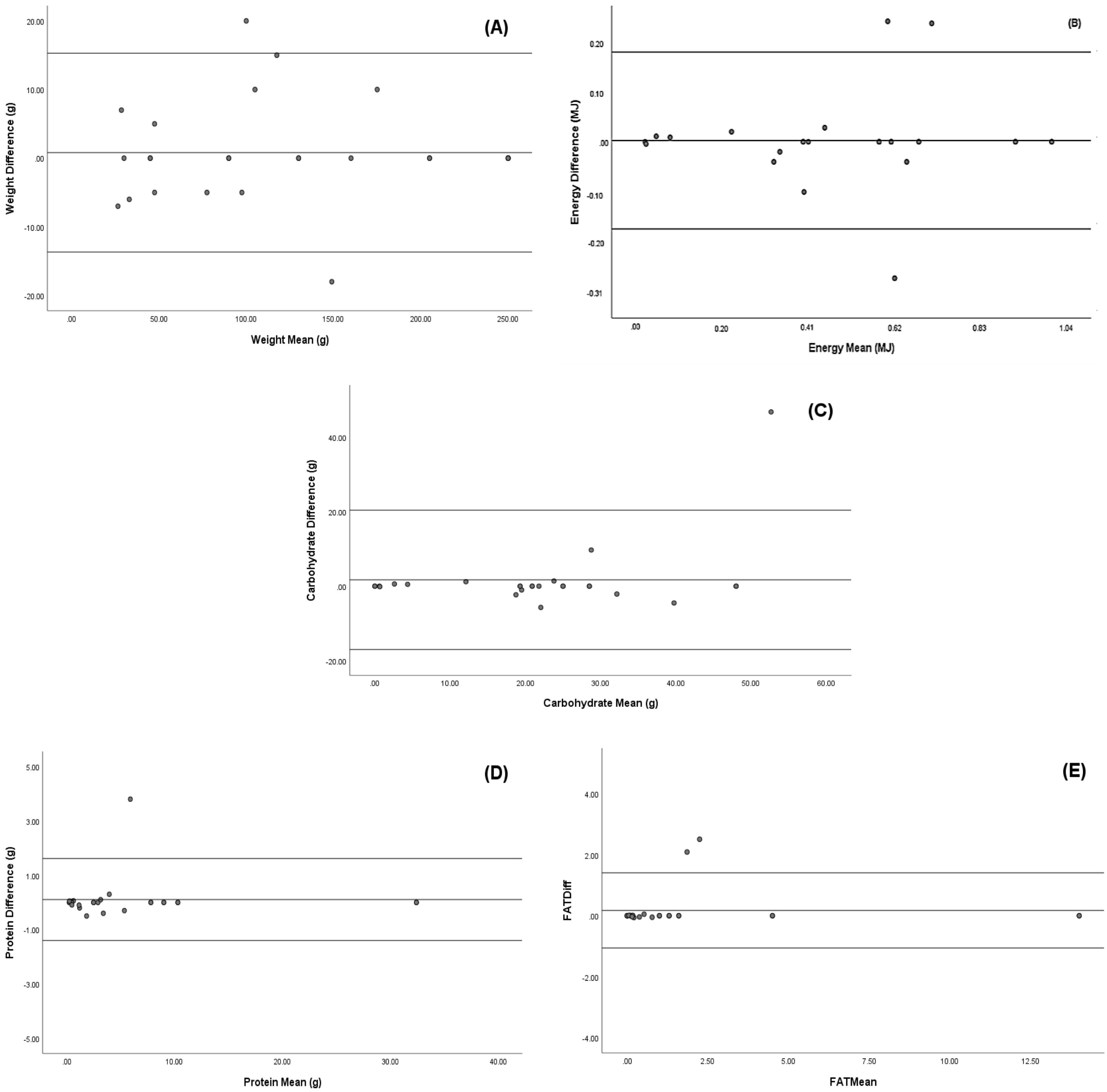

Figure 2: Bland-Altman scatterplots with limits of agreement (mean bias ± 1.96 SD) for test-retest reliability of single foodstuffs from food photography for Weight (A), Energy (B), Carbohydrate (C), Protein (D) and Fat (E).