JMIR Publications
The Leading eHealth Publisher

Relative validity of the food recording app Libro, adapted for young people: a crossover study

Melissa Basso, Liangzi Zhang, George Savva, Kathrin Cohen Kadosh, Maria Traka

Submitted to: Journal of Medical Internet Research
on: June 07, 2024

Preprint
JMIR Publications

Disclaimer: © The authors. All rights reserved. This is a privileged document currently under peer-review/community review. Authors have provided JMIR Publications with an exclusive license to publish this preprint on it's website for review purposes only. While the final peer-reviewed paper may be licensed under a CC BY license on publication, at this stage authors and publisher expressively prohibit redistribution of this draft paper other than for review purposes.

Table of Contents

Preprint JMIR Publications

Relative validity of the food recording app Libro, adapted for young people: a crossover study

Melissa Basso[1]; Liangzi Zhang[2]; George Savva[2, 3]; Kathrin Cohen Kadosh[1]; Maria Traka[2]

[1]School of Psychology Faculty of Health and Medical Sciences University of Surrey Guildford GB
[2]Food and Nutrition National Bioscience Research Infrastructure Quadram Institute Bioscience Norwich GB
[3]Core Science Resources Quadram Institute Bioscience Norwich GB

Corresponding Author:
Maria Traka
Food and Nutrition National Bioscience Research Infrastructure
Quadram Institute Bioscience
Rosalind Franklin Road
Norwich Research Park
Norwich
GB

Preprint JMIR Publications

Abstract

Background: Dietary intake plays a crucial role in health research, yet existing methods for its measurement present several challenges such as participant burden, lengthy recording processes and human errors, and do not account for age-specific variations.

Objective: This study assessed the relative validity of a food recording program within Libro, a real time diet tracking app, compared to a 24-hour recall method, specifically Intake24, among a group of young people sampled from a population vulnerable to eating misbehaviour. The food recording program was customized for the target population based on feedback gathered from a focus group.

Methods: The relative validity of Libro was tested by adopting a cross-over design which recorded food intake over a period of 3 non-consecutive weekdays and 1 weekend day with both methods. The primary outcome was concordance of total energy intake between the two methods, with secondary outcomes focusing on intake of protein, carbohydrates, fats, free sugars, fibre, and trans-fatty acids. Test-retest validity was assessed per each method with the intraclass correlation coefficient; a Bland-Altman plot and t-test were performed to test agreement at group level; correlation coefficient and cross-classification were performed to assess agreement at individual level.

Results: The average intraclass correlation coefficient for energy intake measured by Libro over four days was 0.85 (95% CI: 0.76-0.91). Compared to Intake24, average energy intake recorded using Libro was significantly lower (mean difference: -554 Kcal, 95% CI: -804.1 to -305.6 Kcal, $p < 0.001$), potentially driven by the reduced reporting of foods rich in free sugars. The correlation coefficient for average energy intake measured by Libro vs Intake24 was 0.32 ($p = 0.03$), with only 27.7 % of subjects classified in the same quartile with both methods (? = 0.31, $p = 0.03$). Concordance varied across specific dietary component measures.

Conclusions: While Libro had good test-retest reliability if adopting a multi recall method, it underreported energy and other aspects of dietary intake, along with poor classification performance compared to Intake24. These results might be partially explained by the absence of pop-up memory prompts (which are a feature of Intake24), and the inability to customise the notification system throughout the day. We also suggest that a user-centred interface, graphic design, and implementation of strategies to enhance motivation may play a role in affecting participants compliance and should be improved to enhance the overall food recording process.

(JMIR Preprints 07/06/2024:62707)
DOI: https://doi.org/10.2196/preprints.62707

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✔ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✔ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Preprint JMIR Publications

Original Manuscript

Preprint
JMIR Publications

Type of article:
Original paper

Authors:

Melissa Basso[1], Liangzi Zhang[2], George M Savva[2, 3], Kathrin Cohen Kadosh[1], Maria Traka[2]*

Affiliations:

[1] School of Psychology, Faculty of Health and Medical Sciences, University of Surrey, Guildford, Surrey GU2 7XH, UK
[2] Food and Nutrition National Bioscience Research Infrastructure, Quadram Institute Bioscience, Norwich Research Park, Rosalind Franklin Rd, Norwich NR4 7UQ, United Kingdom
[3] Core Science Resources, Quadram Institute Bioscience, Norwich Research Park, Rosalind Franklin Rd, Norwich NR4 7UQ, United Kingdom

*Corresponding author: maria.traka@quadram.ac.uk

Relative validity of the food recording app Libro, adapted for young people: a crossover study

Abstract:

Background

Dietary intake plays a crucial role in health research, yet existing methods for its measurement present several challenges such as participant burden, lengthy recording processes and human errors, and do not account for age-specific variations.

Objectives

This study assessed the relative validity of a food recording program within Libro, a real time diet tracking app, compared to a 24-hour recall method, specifically Intake24, among a group of young people sampled from a population vulnerable to eating misbehaviour. The food recording program was customized for the target population based on feedback gathered from a focus group.

Methods

The relative validity of Libro was tested by adopting a cross-over design which recorded food intake over a period of 3 non-consecutive weekdays and 1 weekend day with both methods. The primary outcome was concordance of total energy intake between the two methods, with secondary outcomes focusing on intake of protein, carbohydrates, fats, free sugars, fibre, and trans-fatty acids. Test-retest validity was assessed per each method with the intraclass correlation coefficient; a Bland-Altman plot and t-test were performed to test agreement at group level; correlation coefficient and cross-classification were performed to assess agreement at individual level.

Results

The average intraclass correlation coefficient for energy intake measured by Libro over four days was 0.85 (95% CI: 0.76-0.91). Compared to Intake24, average energy intake recorded using Libro was significantly lower (mean difference: -554 Kcal, 95% CI: -804.1 to -305.6 Kcal, $p < 0.001$), potentially driven by the reduced reporting of foods rich in free sugars. The correlation coefficient for average energy intake measured by Libro vs Intake24 was 0.32 ($p = 0.03$), with only 27.7 % of subjects classified in the same quartile with both methods ($\kappa = 0.31$, $p = 0.03$). Concordance varied across specific dietary component measures.

Conclusions

While Libro had good test-retest reliability if adopting a multi recall method, it underreported energy and other aspects of dietary intake, along with poor classification performance compared to Intake24. These results might be partially explained by the absence of pop-up memory prompts (which are a feature of Intake24), and the inability to customise the notification system throughout the day. We also suggest that a user-centred interface, graphic design, and implementation of strategies to enhance motivation may play a role in affecting participants compliance and should be improved to enhance the overall food recording process.

Keywords:

Food recording, dietary intake, nutrition, young people, eating disorders, dietary assessment.

Introduction

Measurements of food choices and dietary intake are central to much health- and diet-related research. Current gold standard methods include recovery biomarkers, which are objective measures of nutrient status. However, they are expensive, do not capture the full complexity of dietary behaviour, and only exist for a subset of nutrients (1). Self-reported dietary assessment methods are thus more commonly used and include food frequency questionnaires (FFQs), 24-hour recalls (24HRs), and food recording diaries (FRs). FFQs measure dietary intake over a longer time, usually 6 to 12 months, categorizing similar food items and quantifying their frequency of consumption. In contrast, 24HRs assess the detail of diet during the previous 24 hours. This makes the instrument more precise in nutrient intake calculation, though more susceptible to day-to-day variation which can be accounted for by the collection of multiple, non-consecutive recalls. FR methods also assess diet over a shorter period, yet the recording process is done in real-time, potentially correcting for retrospective memory bias, but subject to reactivity.

Despite the higher precision of 24HRs and FRs methods compared to FFQs, they still face several challenges such as i) random recording errors; ii) person-specific biases related to e.g., motivation, age, gender; iii) high participant burden for data entry, cognitive load, portion size estimation, and memory; iv) social desirability of food and psychological vulnerability that could trigger underreporting or eating patterns alteration (2, 3). Adolescents and young people may face additional hurdles due to sporadic and social eating, and heightened risk for the development of maladaptive eating behaviour (4, 5). While random errors can be mitigated through repeated measures, systematic errors consistently deviate from true values and may be method specific, leading to biases in results (3). To ensure good reliability and accuracy, new methods need to be validated against existing ones. When a direct comparison to unbiased gold standard methods is not feasible, validation studies should select a reference method that is sensitive to different types of errors and has previously established a certain degree of validity (6). Accordingly, these studies measure relative as opposed to absolute validity, a factor that should be considered in results interpretation. It is also important that

validation studies are conducted across different demographic groups acknowledging age-specific variations that could occur in energy demands, eating behaviour, and self-reporting accuracy (7, 8).
Technology-based tools, particularly in a mobile format, are preferred by young people (9-12). Digital tools ease dietary data collection through e.g., multiple entry modalities, drop-down menus, and more automatised matching procedures to food databases (13). Image-recognition methods are also available, although they are still in their infancy and are not acceptably precise (14). Several studies have validated digital FRs and compared them against 24HRs and found a poor performance for energy intake, and mixed results for nutrients intake (15, 16). However, neither study accounted for age or demographic group as previously recommended. Also, both studies employed a 24HRs interview as reference method, conducted on the same day and meal as the FR log. This experimental set-up might skew results by memory reinforcement processes and prompts of commonly forgettable food items during the interview, unlike the food log app evaluated. Technological tools hold significant promise for enhancing the precision of self-reported dietary assessments while reducing costs and participants' burden. They are generally well-accepted and show good compliance, although adherence may decrease over prolonged recording periods (9, 15). Altogether, this underscores the need of refining these tools on preferences from the target group, and conducting validation studies that address gaps and limitations identified in the literature.
The aim of the present study is to test the adherence to and relative validity of a mobile-based FR app – the Libro app – against a web-based 24HRs i.e., Intake24 in a group of young people using a cross-over design. Given the heightened risk and prevalence of disordered eating behaviour in the target age group, we focused on populations known to be vulnerable (4, 5), especially as vulnerability may not always be apparent and dietary assessment studies should carefully avoid potential triggers. The Libro app (Nutritics Ltd) is equipped with comprehensive food database from different countries, including the official UK national databases i.e., 2021 McCance & Widdowson Composition of Foods integrated dataset & Quadram labelling dataset (17, 18). Libro was chosen as it offers unique features, such as customizable programs with specific media, parameters, daily reminders, selective trackers and feedback display, and real-time compliance tracking through the Nutritics platform. We then tailored a specific program on young people, upon feedback received from a focus group. Intake24 utilizes the UK Nutrient Databank and was adopted as reference method because : i) its interface was developed for maximum engagement of adolescents and young people (19); ii) its relative validity was tested across various populations, including young people (20-22); iii) it is sensitive to a different type of memory errors: while Libro is particularly prone to prospective memory errors (forgetting to initiate a record), Intake24 relies on retrospective memory (recalling what was eaten).

Methods

Focus group and food recording programme

Three young people were recruited, upon given consent, through the McPin foundation to participate in the focus group in March 2023, which was conducted remotely via Microsoft Teams. The McPin foundation is a mental health research charity, and it was chosen to facilitate access to young people with lived experience of psychiatric disorders including eating misbehaviour. Individuals were considered eligible if aged between 14-27 years old and currently living the United Kingdom. Participants were selected to include individuals with previous experience with any digital FR, and at least one person with history of eating disorders. Due to the sensitive nature of the conversation, no camera or voice recordings were used. Two researchers facilitated the focus group. Written records of the discussion were subsequently organized in major themes encompassing: i) media content including welcoming videos, instructions, and troubleshooting; ii) notification system including frequency and format of reminders; iii) food recording and portion size estimation; iv) trackers and

psychological factors including potential harms; v) preferable support routes to accommodate technical issues.

Validation study

Participants recruitment

Participants were recruited through the McPin foundation and considered eligible if between 14-27 years old and currently living in the United Kingdom. No other eligibility criteria were applied. Each participant was asked to sign a consent form before enrolment and received a monetary reward at study conclusion. Recruitment was conducted via mailout to the McPin foundation register of young people interested in mental health research, and delivery of the study was conducted remotely using Microsoft Teams. The sample size was determined based on the required precision in the correlation between the average measures obtained from the Libro and Intake24 methods. To achieve a 95% confidence interval width between 0.21 in Pearson correlation, based on an anticipated correlation of 0.8, it was necessary to enrol 50 subjects to participate in the study. A total of 49 participants were recruited. The protocol for the validation study and focus group were approved by the Ethics Committee of the University of Surrey (Reference number: FHMS 22-23 099).

Experimental design

The data were collected between May and August 2023, using the cross-over design illustrated in Figure 1. Twenty-seven participants included in the study recorded their food intake using the Libro app on three alternative weekdays (i.e., Monday, Wednesday, and Friday) and one weekend day (i.e., Saturday) during week 1. In the subsequent week (week 2), they followed the same recording protocol using Intake24. Conversely, 22 participants used Intake24 in week 1 followed by Libro in week 2. Participants were assigned to either sequence by alternation. The group size difference was due to a technical issue with the Libro app's initialization, which meant that two participants initially allocated to the Libro-Intake24 sequence were reassigned to the alternate order.

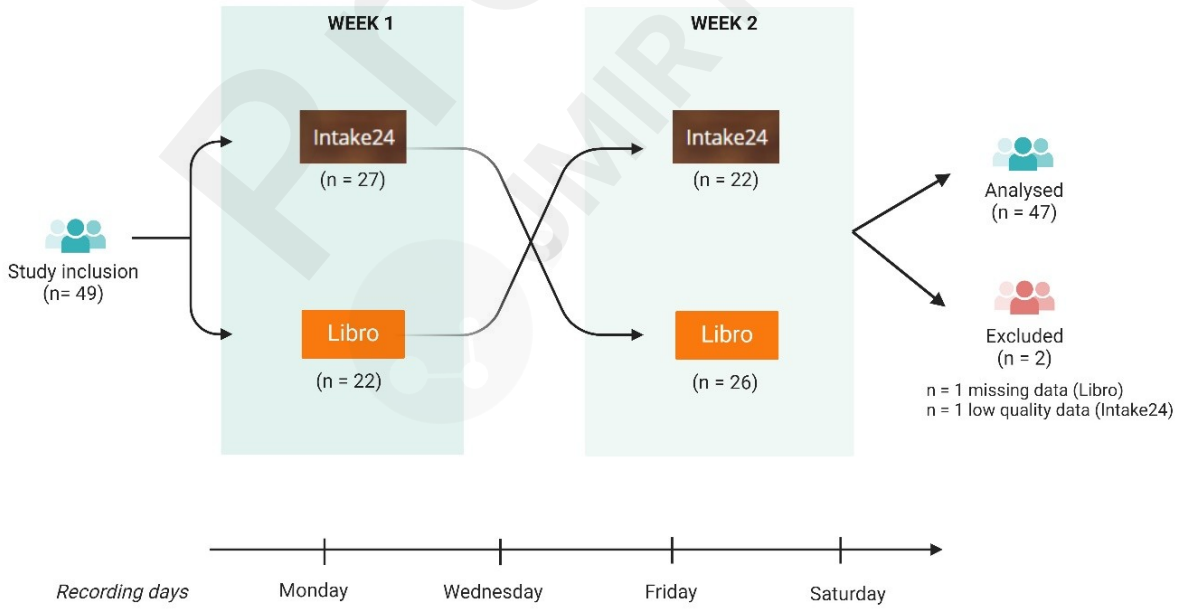


Figure 1. Experimental design and participant flow: Forty-nine participants were included in the study and randomized to each study branch. Twenty-two participants recorded their food intake using the Libro app on three alternative days (i.e., Monday, Wednesday, and Friday) and one weekend day (i.e., Saturday) during week 1. In the subsequent week (week 2), they followed the same recording protocol

using Intake24. Twenty-seven participants first recorded their diet with Intake24, and subsequently with Libro. One participant failed to record their diet with Libro and was thus excluded from the analysis. One further participant was excluded due to implausible reporting of energy intake higher than 5000 Kcal with Intake24. Created with BioRender.com.

Use of Libro and Intake24

An initial online appointment was arranged with each participant to deliver verbal instructions for both Libro and Intake24. Each participant then received an anonymised code and email address for registration to the Nutritics study account and log in to Libro. Manual emails were sent a couple of days before the first food log with both methods. A day before initiating the Libro program, participants were also notified via a welcoming message and an instructional video within the app. For Intake24, automatic reminders were also sent through the app on the day before recording their diet. To ensure the timely submission of daily food records, real-time monitoring was conducted by the lead researcher through the analytics platform of Nutritics and Intake24 to track user activities. Reminders via email were sent if food records were missing.

Measures

To evaluate the validity of the Libro food record, dietary outcomes captured by Libro and Intake24 were compared. The primary outcome focused on total daily energy intake, while secondary outcomes were intakes of carbohydrates, fats, protein, fibres, free sugars (i.e., non-milk extrinsic sugars) and trans-fatty acids intake. The selection of carbohydrates, fats, and protein as secondary outcomes align with other nutritional studies; whereas fibres, trans-fatty acids and free sugars were chosen based on their favourable (fibres only) and unfavourable effects on health and increasing research looking at diets high in such nutrients (23-25).

Data processing and statistical analysis

Adherence

Protocol adherence was quantified by the number of daily single recalls (SRs) completed, intended as record of foods and drinks on a single given day, done by each subject, with each given method. Full adherence was intended as completion of all four assigned SRs, difference in full adherence between methods was assessed with the McNamar Chi-squared test and considered significant with p value < 0.05. Subjects that did not complete any SR with either Libro or Intake24 were excluded from the subsequent analysis.

Data quality check and data cleaning

Data quality checks of SRs were carried out through data visualization with boxplots at the end of data collection. Recorded portion sizes of outlying points were checked to explore potential recording errors. When possible, participants were also consulted when recording errors were suspected. Finally, we quantified underreporting within each method by calculating the percentage of SRs showing energy intake lower than 400 Kcal. To maximise data quality and minimise the impact of random recording errors, SRs reporting an energy intake (EI) higher than 5000 Kcal were excluded. We have compared results obtained with and without filtering high values as a sensitivity analysis and no differences in trends and effects were detected. No filters were applied for SRs lower than 400 Kcal.

Statistical analysis

Single recalls and test-retest reliability

The day-to-day variation in total energy intake from single recalls per participant was quantified as

the difference between the highest and the lowest SR and summarized across participants as median and interquartile range (IQR) for each method, with a Wilcox test assessing significant differences between methods. The intraclass correlation coefficient (ICC) was also used to assess test-retest reliability within methods and computed following a two-way mixed-effects model, assessing the reliability when the average of four (rather than single) measurements is used. Correlation coefficients were also calculated between single entries within each method to facilitate distinction between method-related differences in individual-level assessment and normal daily variation.

Individual and group agreement between methods

We first calculated the mean average EI and nutrient intake (NI) across SRs as recorded with Libro and Intake24 separately for each participant. A paired t-test and Bland-Altman plot were used to assess agreement at group level and presence/direction of difference in average intakes between methods. For the Bland-Altman plot, individual EI was represented by the median of SRs, as the distribution of difference values using mean of SRs, including log-transformed data, did not conform to the normality assumptions. A mixed-effects model was then performed to confirm the difference in average EI between methods while also controlling for potential order effects from the study. A correlation coefficient was determined to evaluate agreement between methods on an individual level using a continuous scale. To adjust for variations in test-retest reliability across methods, a deattenuated coefficient was computed following the approach outlined by Trafimow et al. (26). Cross-classification analysis was conducted to assess agreement at the individual level on a categorical scale, utilizing quantiles. A favourable outcome was defined as more than 50% of individuals being classified into the same category, with fewer than 10% being classified into diametrically opposite categories. Finally, kappa statistics was computed to quantify inter-rater reliability on this categorical scale.

The choice of statistical tests to performed was guided by a review from Lombard and colleagues (27). Before each statistical test, normality assumption was assessed with qq plots. When the normality assumption was met, parametric tests were performed, and mean values used; when data were not normally distributed, non-parametric alternatives were performed, and median values used. A p-value of 0.05 was adopted as significant thresholds for hypothesis tests. Average NI was first summarised as energy density for visual comparison with reference values from the National Diet and Nutrition Survey (NDNS). For analysis, average NI per participant was adjusted based on the residual model where the dependent variable is absolute nutrient intake in grams, the independent variable is total EI (28). When applying the residual model, five values (within free sugars and trans-fatty acids intake measures) resulted just below zero and were thus considered as 0 for data interpretability. Data were pre-processed and analysed using the R language in RStudio (version 4.3.0) using the packages: “networkD3 0.4”, “car 3.1-2”, “lme4 1.1-35.1”, “irr_0.84.1” (29-33). The scripts used for data pre-processing and analysis are available on GitHub (34).

Results

Results from the focus group and FR program creation

The focus group included three females aged 23, 25, and 26 years old who had previous experience with FR apps; two had history of eating disorders. The consultation highlighted the following points:

1. It would be helpful to include FR instructions both in a written and video format, as well as examples for troubleshooting.
2. Notifications should be sent throughout the day to aid food recording and commonly forgotten ingredients e.g., sugar and sauces. Where possible, the notifications should be customized to ones’ eating schedule as “*useless and potentially annoying*” otherwise.
3. Young people find the food entry process tedious and suggest employing multiple entry modalities, as well as AI-based picture recognition and portion size estimation to aid the process

and limit the FR-related burden. Where not possible, they suggest including examples of portion sizes and highlight the importance of an easy-to-use graphic interface.
4. Young people confirm that feedback during the FR process could potentially alter their eating patterns and should therefore be limited to the period after study completion. The young people also pointed out that trackers could trigger obsessive thoughts and disordered eating behaviour, and they suggest phrasing reminders in a “neutral and gentle” format.
5. The availability of virtual support would make young people more comfortable with the process and they highlight that it could help avoid issues with data quality and missing entries.

Based on the insights gain from the focus group, the Libro program was customized as follows (see Figure S1):
1. A welcoming message was included and delivered on the day before the start of the study. Instructions were directly accessible through the app, in both written and video format.
2. A notification system was created which sent 4-5 reminders within the app throughout the day and notes for commonly forgotten ingredients were included. Note that individual level customization of notifications and pop-up memory prompts were not possible.
3. Changes at the user-interface level and inclusion of AI-based recognition functions were not possible. However, the Libro app allows for multiple entry modalities i.e., via text, barcode, and voice notes.
4. No trackers were shown during the study or delivered after completion. Reminders and prompts were formulated neutrally.
5. The email address of the researcher responsible for data collection was made available within the program for any queries.

Example of the food recording process and graphic layers can be found in Figure S2.

Results from the validation study

Participants

Of 49 participants included into the study, one didn’t complete any SR in Libro, and one was excluded due to poor data quality, resulting in 47 participants for analysis. Fourteen participants were female (29.8%), mean age (+- standard deviation) was 23.2 (± 2.4) and mean BMI was 24.3 (±7.4) (see Table 1). Five participants had past or current diagnosis of a psychiatric disorder (two participants had eating disorders, three anxiety and/or depression).

Table 1. Demographics of the included sample

	Final sample	**Males**	**Females**
	n = 47	n = 33 (70.2%)	n = 14 (29.8%)
mean age	23.2 (± 2.4)	23.6 (± 1.7)	22.8 (± 2.8)
age range	19.1 - 27.1	20 - 27	19.1 - 27.1
weight (Kg)	63 (± 9.6)	65.2 (± 7.9)	57.6 (± 11.4)
height (cm)	163.65 (± 17.5)	164.6 (± 19.8)	161.3 (± 10.2)
BMI	24.3 (±7.4)	25.2 (± 8.3)	22.2 (± 4.4)

±: standard deviation

Adherence

Of the 49 participants included in the study, 46 participants (94%) completed 4 SRs with Intake24,

the remaining 3 (6%) completed 3 SRs. Forty-three participants (88%) completed 4 SRs with Libro, 4 participants (8%) completed 3 SRs, 1 participant complete 2 SRs, 1 completed none. Difference in full adherence did not differ significantly between the two methods (McNemar Chi-squared test, p value = 0.45).

Data quality check and data cleaning

Figure 2 shows the daily EI from each single recall by each given day. Daily EI captured by Intake24 showed a higher number of outliers with a trend towards the higher end. Manual inspection of the recorded items found a few implausible portion sizes underlying these records. Single recalls higher than 5000 Kcal were then removed for both methods from further analysis. Data quality was also inspected in terms of underreporting defined as EI < 400 Kcal. Three SRs (1.6%) were identified for Intake24, although no participant had an average EI below 400 Kcal when considering the mean of their recorded SRs. Ten SRs (5.3%) recorded using Libro were lower than 400 Kcal, with two participants (4.1%) reporting less than 400Kcal on average across all their SRs.

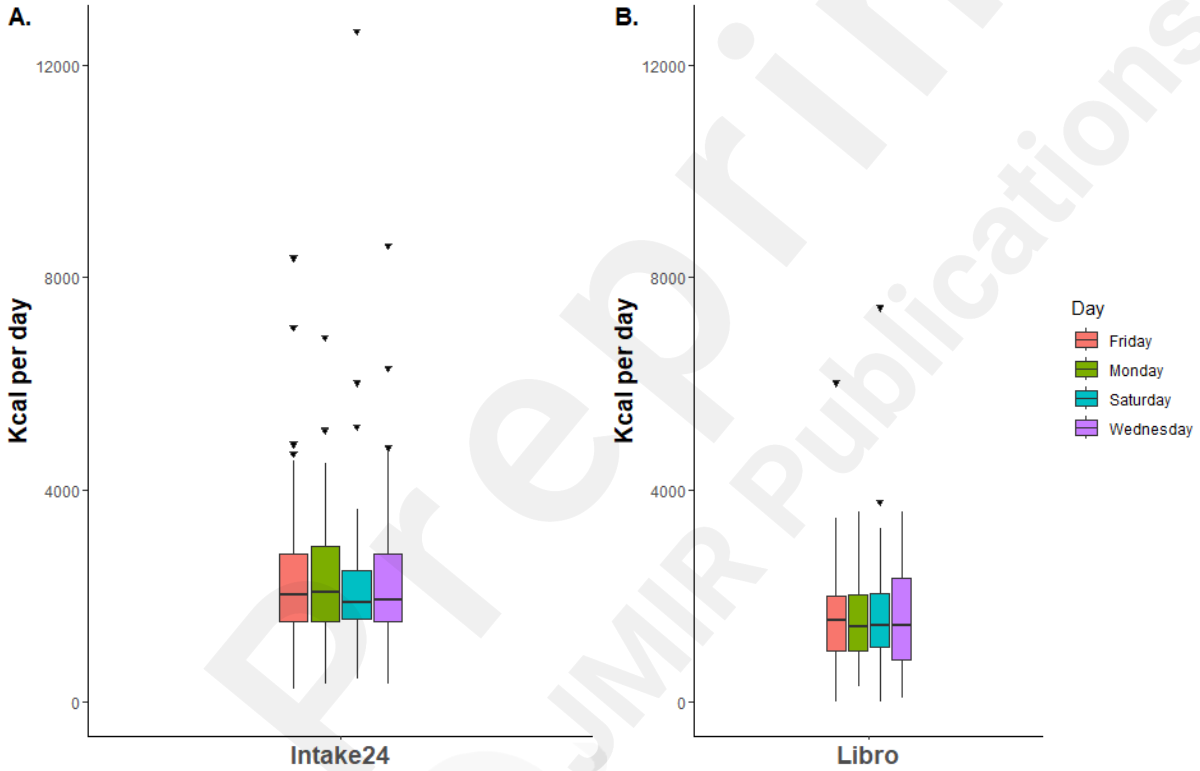


Figure 2. The distribution of daily single recalls of total energy intake by day from all participants recorded with Intake24 (left) and Libro (right).

Results on energy intake

Single recalls and test-retest reliability

Day to day variations of SRs per participant can be seen in Figure 3. The median of EI variation was 975.9 Kcal (IQR = 583.1 – 1577.4 Kcal) for Intake24 (in Figure 3A), 754.8 Kcal (IQR = 539.7 – 1360.4 Kcal) for Libro (in Figure 3B) (paired sign-rank Wilcox test: p value > 0.05). The ICC was similar between methods: the average ICC over four days for Intake24 was 0.83 (95% CI: 0.73 – 0.90), ICC for Libro was 0.85 (95% CI: 0.76- 0.91). Correlation between single entries were moderate within both methods, ranging from a Spearman correlation coefficient of 0.48 to 0.64 for Libro (all p's ≤ 0.001), from 0.55 to 0.71 for Intake24 (all p's ≤ 0.0001).

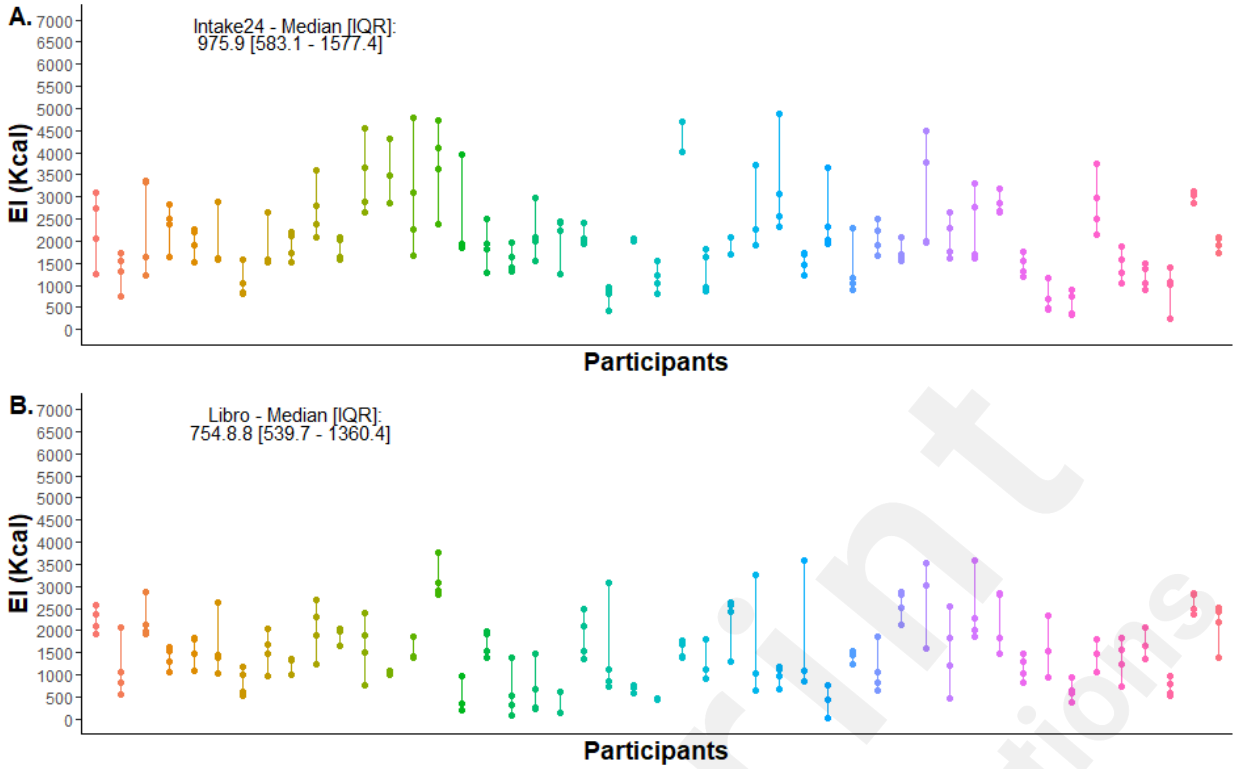


Figure 3. Day-to-day variations of the energy intake for each participant captured using Intake24 (A) and Libro (B). Different colours refer to different participants. EI: energy intake, IQR: interquartile range.

Individual and group agreement between methods

The mean average daily EI across participants captured by Libro was 1512 Kcal, the median was 1542 Kcal, minimum and maximum values were 306 and 3138 Kcal. The average EI captured by Intake24 was 2094 Kcal, the median was 2031 Kcal, minimum and maximum values were 592 and 4351 Kcal. Qq plots for energy intake analysis can be seen in Figures S3-S5. The Bland-Altman plot estimated a bias of -554.9 Kcal (95% CI: -804.1 to -305.6 Kcal) with 95% limits of agreement being -2263.5 to 1153.7 Kcal (see Figure 4A). Visual inspection of the Bland-Altman plot shows no dependency of differences on average EI values. Agreement at group level was also assessed using a paired t-test which revealed a significant difference in EI assessment ($p < 0.001$) as shown in Figure 4B. Analysis through linear mixed models confirmed a significant impact of the assessment method on EI variation ($p = 3.45e-05$) while the sequence of method application did not significantly influence the results ($p = 0.147$). Analysis of individual agreement between average SRs revealed a Pearson correlation coefficient between methods of 0.32 ($p = 0.03$) as illustrated in Figure 5A. After adjusting for the attenuation effect, the correlation coefficient rose to 0.38. Utilising a cross-classification approach as depicted in Figure 5B, 27.7% of the subjects were categorised in the same quartile, 4.3% were placed in opposite categories. To further evaluate inter-rater reliability on a categorical scale, weighted Kappa statistics was employed indicating a fair agreement ($\kappa = 0.31$, $p = 0.03$)

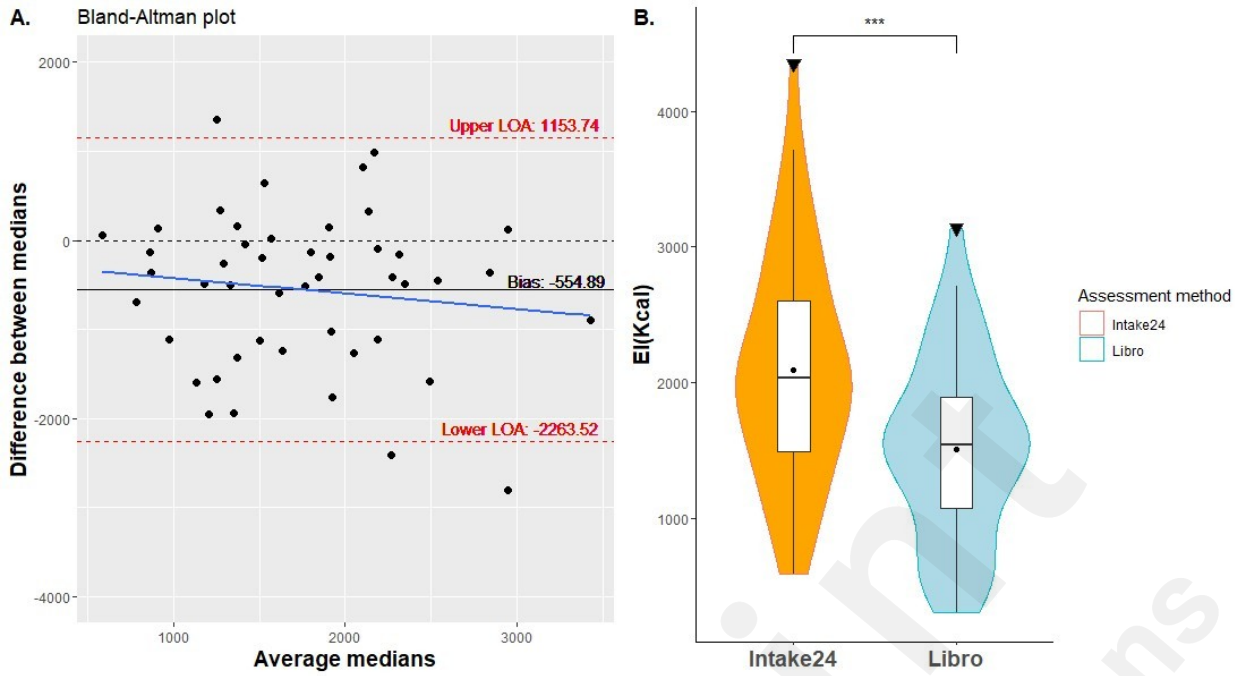


Figure 4. **Group-level analysis. (A.) Bland-Altman plot.** The differences between medians of SRs are plotted against the averages of the medians, with limits of agreement indicated by the dashed red lines and bias by the continuous black line. The blue line is used to investigate if any trend in the differences across the range of measurements exist. **(B.) Violin and boxplots**. Violin and boxplots of Energy Intake (EI) in Kcal as recorded by Intake24 (orange) and Libro (light blue). Triangles represent outliers. ***: p value < 0.001.

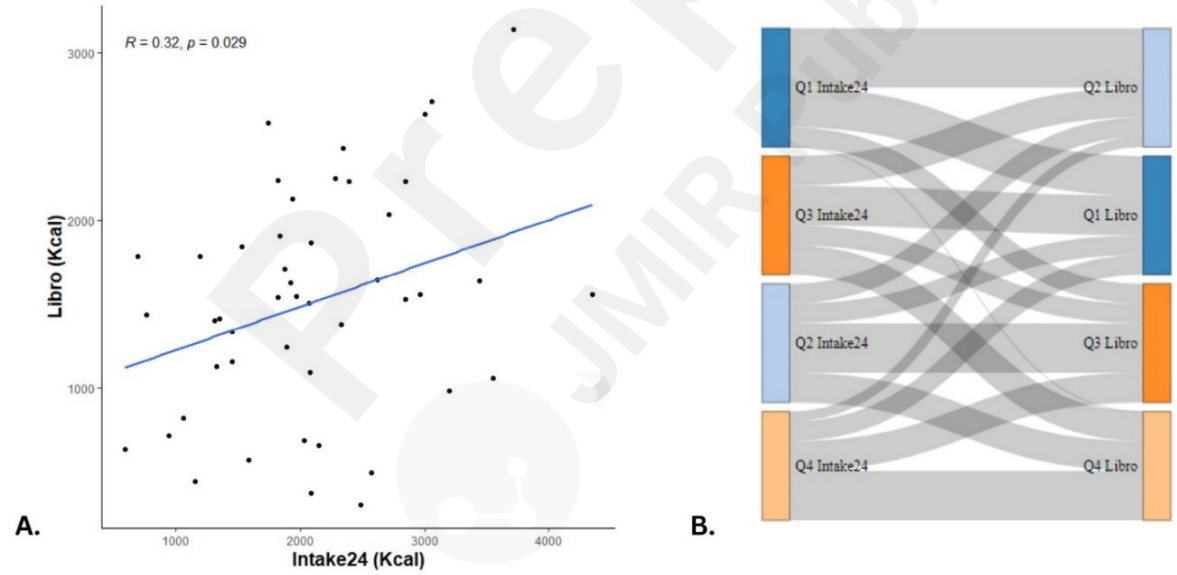


Figure 5. Individual level analysis. (A.) Scatter plot of average individual energy intake (EI) values in Libro versus Intake24 with a linear regression line**. (B.) Sankey diagram** depicting participant flow from one quartile as assessed by Intake24 (left) to the same or different quartile as assessed by Libro (right).

Results on nutrients intake

Summary statistics (mean and median intake) of NI can be found in Table 2, along with standard scores reported by the NDNS for the 19-64 age group. The NI is expressed in nutrient density (%) for visual comparability. The percentage of energy derived from fats, carbohydrates, and protein

intake was comparable to the NDNS data for both Libro and Intake24. Free sugar intake was comparable for Intake24 only, whereas trans-fatty acid intake was comparable to NDNS data for Libro only.

We then applied the individual and group level analysis to energy-adjusted NI (Table 3). Given the non-normal distribution of NI data, non-parametric methods were employed. Spearman correlation coefficients between methods ranged from -0.04 (for fats) to 0.27 (for protein) with non-significant p values (all r's $\leq$ 0.27; all p's $\geq$ 0.07) (see Figure S6). Assessment bias was quantified as median differences and ranged from -0.42 g (for trans-fatty acids) to -93.93 g (for carbohydrates) and Wilcoxon signed rank test showed a p value lower than 0.05 for all analysed nutrients, suggesting that Libro significantly underestimate NI compared to Intake24. The biggest difference was seen for free sugars and trans-fatty acids, with a 67% and 47% underreporting when using Libro, respectively. Cross-classification based on quartiles showed that the percentage of subjects that were correctly classified ranged between 12.77 % (for fats) and 38.3 % (for proteins and trans-fatty acids). Between 8.51% (for protein, fibres, free sugars, and trans-fatty acids) and 12.77% (for carbohydrates) of the participants fall within opposite categories. Weighted Kappa statistics ranged from -0.09 (for fats) and 0.29 (for proteins) with p values of 0.05 suggesting a significant correlation between methods only for protein.

Preprint JMIR Publications

Table 2. Summary statistics of energy intake (expressed Kcal) and nutrients intake (expressed in nutrient density) as assessed by Intake24 and Libro.

	Intake 24		Libro		NDNS[a]	
	Mean (sd)	Median (lower-upper 2.5 percentile)	Mean (sd)	Median (lower-upper 2.5 percentile)	Mean (sd)	Median (lower-upper 2.5 percentile)
Energy (Kcal)	2094.85 (828.08)	2031.92 (704.59-3691.13)	1512.02 (673.52)	1542.61 (386.97-2698.34)	1882 (628)	1815 (864-3176)
Protein (%)	17.1 (3.96)	17.02 (9.74-24.93)	20.83 (7.94)	19.41 (11.45-41.91)	16.5 (4.2)	16.0 (10.3-25.6)
Fat (%)	31.95 (5.55)	33.06 (22.49-39.93)	34.02 (6.4)	33.77 (24.44-47.1)	32.9 (6.6)	33.4 (18.9-44.9)
Carbohydrate (%)	49.84 (7.46)	49.06 (37.53-66.43)	43.34 (10.75)	45.78 (11.63-58.83)	45.5 (7.7)	45.5 (30.0-60.5)
Fibers (%)	1.49 (0.49)	1.37 (0.97-2.89)	2.31 (0.88)	2.29 (0.95-3.91)	-	-
Free sugars (%)	11.8 (5.78)	11.28 (4.62-22.17)	5.36 (4.06)	4.35 (0.02-14.58)	11.6 (6.2)	10.7 (2.4-25.0)
Trans Fatty-Acids (%)	0.37 (0.12)	0.37 (0.18-0.65)	0.67 (0.53)	1.5 (0.14-2.17)	0.7 (0.3)	0.6 (0.2-1.3)

[a]Mean and median intake from the National Diet and Nutrition Survey (NDND) for the years 19-64 age groups for reference

Sd: standard deviation

Table 3. Individual and group level analysis of energy-adjusted nutrients intake between Intake 24 and Libro.

	Intake24 median (IQR)	Libro median (IQR)	Spearman (r)	Spearman (p value)	Median difference (95% CI)	Wilcoxon signed rank (p value)	Cross-classification % same quartiles	Cross-classification % opposite quartiles	Weighted Kappa stat (value)	Weighted Kappa stat (p value)
Protein(g)	90.25(76.05-99.79)	67.52(59.93-82.6)	0.27	0.07	-17.88 (-23.45,-13.91)	3.79972491e-07	38.3	8.51	0.29	0.05
Fat(g)	76.29(67.77-81.51)	56.47(51.87-61.69)	-0.04	0.77	-20.61 (-26.81,-13.79)	3.438657e-09	12.77	10.64	-0.09	0.53
Carbohydrate(g)	273.71(253.75-293.78)	179.78(162.9-192.15)	0.21	0.16	-93.93 (-107.95,-83.56)	2.8e-14	34.04	12.77	0.19	0.2
Fibers(g)	14.35(12.55-16.22)	16.58(13.29-20.22)	0.2	0.18	2.07 (0.51, 3.49)	0.00938259422864	34.04	8.51	0.22	0.13
Free sugars(g)	61.69(42.18-NA)	20.26(10.8-30.37)	0.21	0.15	-41.39 (-49.9,-32.07)	1.5191e-11	31.91	8.51	0.19	0.2
Trans Fatty-Acids(g)	0.88(0.76-0.98)	0.45(0.26-0.6)	0.16	0.27	-0.42 (-0.56,-0.34)	6.38523843 86e-05	38.3	8.51	0.19	0.2

IQR: Interquartile Range, CI: Confidence Interval

Preprint JMIR Publications

Discussion

The present study customised and tested the validity of a food recording program – built in the Libro app – against a 24-hour recall system, Intake24, in a group of young people sampled from a population vulnerable to eating misbehaviour. We found that both Libro and Intake24 presented a comparable day-to-day variation and good test-retest reliability. However, Libro was poorly correlated with Intake24 even with four measurements per method, and appeared to underestimate both energy and macronutrients intake, with an underreporting bias for EI of 27% (554 Kcal) and wide limits of agreement (-1153.7 to 2263.5 Kcal). These results are consistent with another study where a FR app underestimated energy intake – although of 6% – and about half of the nutrients relative to a 24-hour recall (16). Specifically, Ocke et al. found the FR app to underestimate the total intake in saturated fats, carbohydrates, mono- and disaccharides as expressed in grams per day; but not others such as proteins (g and Energy %), sodium (mg), saturated fats, carbohydrates and modo- and disaccharides as expressed in energy percentage (16). Other authors focused on the accuracy of dietary recording of a controlled meal and found that participants significantly underreported their energy intake when using a FR app with wide limits of agreement. With a 24HRs method, participants also underreported energy intake, albeit not significantly. Worth noting, the FR app was generally described as easy to use and likely to be used daily (15). Libro reported 67% less free sugar intake when compared to Intake24, and a considerable underreporting of 47% is also seen for trans-fatty acids. Few validation studies have concentrated on free sugars and trans-fatty acids, despite their significant health implications. Recognizing this gap, we included these nutrients in our assessment, influenced by previously established accuracy of Intake24 i.e., findings that mean free sugar intake reported by Intake24 is within 3% of the interviewer-led recalls (20). Our results suggest that Libro might lack accuracy in capturing free sugars and trans-fatty acids intake, although future studies should perform reliability comparisons based on number of single recalls. Given the high day-to-day variation of such nutrients, a higher number of single recalls might indeed be needed to capture true intake.

Prior work suggests that the underreporting bias observed with the Libro app might be attributable to the absence of memory prompts for commonly forgotten ingredients, a feature that Ocke and colleagues recommend including in self-reported dietary assessment tools. (16). The absence of tailored pop-up probs in Libro, unlike Intake24, could also explain the relative underreporting of free sugars by increasing omission of sugar-rich items such as sugar tablets, sauces, sweetened beverages, and snacks for recording. A comparison of nutrient density between Libro and Intake24 against NDNS data showed visually comparable values for protein, fats, carbohydrates, and trans-fatty acids. This might suggest that Libro might effectively gauge the proportion of NI, despite the general tendency to underestimate overall intake. However, the reported consumption of free sugars in Libro was still notably lower compared to the NDNS data, highlighting the need for further refinement in Libro's methodology to enhance the accuracy of free sugar intake estimates.

We also evaluated agreement at the individual level, using continuous and categorical scales and observed poor performance in both scenarios. Our findings revealed that only 27.7% of subjects were accurately ranked in their respective quartiles for EI, with 4.3% of subjects placed in opposite categories depending on method, although this might also reflect genuine day-to-day or week-to-week variation in intake. Both correlation and Kappa coefficients showed a fair and significant agreement around 0.30, and similar trends were observed for NI estimation, suggesting a high rate of measurement error. To disentangle discrepancies in individual-level accuracy and variations from daily fluctuations in EI, we calculated

correlation coefficients for single entries within each method and observed overall moderate levels of agreement ranging from 0.48 to 0.71, consistent with the reported high ICC for average unit .These results suggest that while Libro may be fairly consistent within its own measures, and present good test-retest reliability when adopting a multiple recall method, it may not accurately capture individual dietary behaviours compared to Intake24.

Self-reported dietary assessment methods are sensitive to participant compliance, and they depend heavily on motivational processes and user experience. To address this, we co-produced the customization of the food recording program together with a small group of young people. However, despite implementing several requested features (e.g., media, instructions, notifications, easily accessible support), we were not able to execute changes at the user interface, graphic and software level. Participants, including young people, seem to prioritize the user-interface layout and are more motivated by colourful and brighter icons (35, 36). An easy-to-use interface with a minimal number of layers is also preferred (37). Therefore, it is plausible that the lack of a user-centric design, coupled with the slower processing speed of the Libro app as compared to Intake24, and unfamiliarity with the FR process could have contributed to lower user engagement. Consistent with that, a study found that experienced app users tend to underestimate intake the least (11). These factors may have led to selective and under-reporting with Libro as users might have sought to simplify the food recording process or forgot to log items. Similarly, our inability to customize the notification system might have also played a role. Feedback from our focus group suggested that notifications could be counterproductive if they did not coincide with a participant's individual eating schedule, potentially disrupting the recording accuracy. The integration of gamification strategies that utilize extrinsic rewards could also enhance user engagement in health-focused digital solutions (38). By embedding these strategies within a framework that also emphasizes psychological and intrinsic rewards, the overall experience could become more compelling and effective. This approach may ensure that while extrinsic rewards such as points or badges motivate initial user interest and participation, enhancing intrinsic motivation promotes e.g., self-determination, competence, and a sense of belonging, sustaining long-term engagement and commitment.

Finally, it is important to acknowledge several limitations of the current study. First, we did not collect data on level of experience with FRs, which limits our ability to determine whether varying levels of familiarity with the task influenced data quality. Second, the online nature of this study could have influenced participants' motivation, biasing their commitment throughout the process. While monetary incentives were used to encourage enrolment, we did not employ reward strategies aimed at intrinsic motivation which could have improved compliance. This may be particularly significant for higher-burden tasks, such as completing FRs with the Libro app. Third, Intake24 recorded a higher number of entries with EI exceeding 5000 Kcal, a finding likely attributable to technical errors in portion size recording. We cannot rule out that Intake24 might have presented a degree of overreporting, although this is unlikely for several factors: i) Intake24 has been validated in multiple studies and reported favourable outcomes; for example Bradley et al. compared intake24 validity to an interviewer-led 24HRs and reported that intake24 underestimates energy intake of as little as 3% in 11-16 years old, whereas it shows agreement for 17-24 years old (although data recorded were from the same day which could increase observed validity) (20); ii) we used a multiple-recall protocol, which is designed to minimize or eliminate recording errors; iii) we implemented data filtering processes prior to analysis. Although both Libro and Intake24 are built on the National food composition database, this does not include free sugar and trans-fatty acids intake. Thus, distinct computational methodology employed to gap-fill missing values might have affected the results (39). Finally, we acknowledge that the focus group was constrained by a small sample size and the absence of video recordings. However, it was

crucial for the co-production of the program, for gaining insights into the existing barriers and facilitators associated with food recording tasks and to avoid any trigger of maladaptive eating behaviour. Future studies should consider incorporating a co-production approach including higher sample sizes and employing qualitative analysis tools such as NVivo to enhance data interpretation (40). A qualitative feedback collection after study completion could also be helpful in developing future tools targeted at this population group and in optimising existing features unique to FR such as barcode, voice, and image recognition.

In conclusion, our study demonstrated good test-retest reliability for the Libro app, yet it exhibited suboptimal performance in assessing dietary intake in a population vulnerable to eating misbehaviour. Indeed, Libro tended to underestimate food consumption, likely due to i) the absence of pop-up prompts for commonly forgotten items, ii) the need for consistent user engagement throughout the day without the support of customized notifications, and iii) the lack of reward strategies that address both extrinsic and psychological motivators. Future studies should focus on optimizing the app's graphic design and underlying technology and stratify participants by their experience with FR. In parallel, we suggest considering the psychological aspects of the food recording process and to develop strategies that target motivational factors to improve engagement and accuracy.

Acknowledgment

MB: Conceptualization, Methodology, Investigation, Formal analysis, Visualization, Writing – original draft. LZ: Conceptualization, Methodology, Writing – Review. GS: Methodology, Writing – Review, Supervision. KCK: Conceptualization, Writing – Review, Supervision, Funding acquisition. MT: Conceptualization, Review, Methodology, Supervision, Funding acquisition.

We would also like to acknowledge Nutritics for allowing us to create the customised version of the Libro app. We would like to thank Dr Laura Bardon for providing useful comments during manuscript preparation.

This work was supported by Faculty of Health and Medical Sciences Enhanced Bid support grant (KKC) and by the UKRI Biotechnology and Biological Sciences Research Council (BBSRC) Core Capability Grant BB/CCG2260/1 and its constituent project BBS/E/QU/23NB0006 (Food & Nutrition National Bioscience Research Infrastructure).

All data are available upon request.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Abbreviations

24HRs: 24-hour recall
EI: energy intake
FFQ: food frequency questionnaire
FR: food recording
NDNS: National Diet and Nutrition Survey
NI: nutritional intake

References

1. Jenab M, Slimani N, Bictash M, Ferrari P, Bingham SA. Biomarkers in nutritional epidemiology: applications, needs and new horizons. Hum Genet. 2009;125(5-6):507-25.PMID:19357868
2. Das SK, Miki AJ, Blanchard CM, Sazonov E, Gilhooly CH, Dey S, et al. Perspective: Opportunities and Challenges of Technology Tools in Dietary and Activity Assessment: Bridging Stakeholder Viewpoints. Adv Nutr. 2022;13(1):1-15.PMID:34545392
3. Kirkpatrick SI, Baranowski T, Subar AF, Tooze JA, Frongillo EA. Best Practices for Conducting and Interpreting Studies to Validate Self-Report Dietary Assessment Methods. J Acad Nutr Diet. 2019;119(11):1801-16.PMID:31521583
4. Swanson SA, Crow SJ, Le Grange D, Swendsen J, Merikangas KR. Prevalence and correlates of eating disorders in adolescents. Results from the national comorbidity survey replication adolescent supplement. Arch Gen Psychiatry. 2011;68(7):714-23.PMID:21383252
5. Stice E, Marti CN, Durant S. Risk factors for onset of eating disorders: evidence of multiple risk pathways from an 8-year prospective study. Behav Res Ther. 2011;49(10):622-7.PMID:21764035
6. Zhang L, Misir A, Boshuizen H, Ocke M. A Systematic Review and Meta-Analysis of Validation Studies Performed on Dietary Record Apps. Adv Nutr. 2021;12(6):2321-32.PMID:34019624
7. University of Cambridge NSR. National Diet and Nutrition Survey. [data series]. 7th Release. UK Data Service 2019;SN: 2000033
8. Jones L, Ness A, Emmett P. Misreporting of Energy Intake From Food Records Completed by Adolescents: Associations With Sex, Body Image, Nutrient, and Food Group Intake. Front Nutr. 2021;8:749007.PMID:34966768
9. Eldridge AL, Piernas C, Illner AK, Gibney MJ, Gurinovic MA, de Vries JHM, et al. Evaluation of New Technology-Based Tools for Dietary Intake Assessment-An ILSI Europe Dietary Intake and Exposure Task Force Evaluation. Nutrients. 2018;11(1).PMID:30597864
10. Boushey CJ, Harray AJ, Kerr DA, Schap TE, Paterson S, Aflague T, et al. How Willing Are Adolescents to Record Their Dietary Intake? The Mobile Food Record. JMIR mHealth and uHealth. 2015;3(2):e47.PMID:26024996
11. Boushey CJ, Kerr DA, Wright J, Lutes KD, Ebert DS, Delp EJ. Use of technology in children's dietary assessment. European Journal of Clinical Nutrition. 2009;63(S1):S50-S7.PMID:19190645
12. Casperson SL, Sieling J, Moon J, Johnson L, Roemmich JN, Whigham L. A Mobile Phone Food Record App to Digitally Capture Dietary Intake for Adolescents in a Free-Living Environment: Usability Study. JMIR mHealth and uHealth. 2015;3(1):e30.PMID:25775506
13. Zhao X, Xu X, Li X, He X, Yang Y, Zhu S. Emerging trends of technology-based dietary assessment: a perspective study. Eur J Clin Nutr. 2021;75(4):582-7.PMID:3082535
14. Hochsmann C, Martin CK. Review of the validity and feasibility of image-assisted methods for dietary assessment. Int J Obes (Lond). 2020;44(12):2358-71.PMID:33033394
15. Chen YS, Wong JE, Ayob AF, Othman NE, Poh BK. Can Malaysian

Young Adults Report Dietary Intake Using a Food Diary Mobile Application? A Pilot Study on Acceptability and Compliance. Nutrients. 2017;9(1).PMID:28098770
16. Ocke M, Dinnissen C, Stafleu A, de Vries J, van Rossum C. Relative Validity of MijnEetmeter: A Food Diary App for Self-Monitoring of Dietary Intake. Nutrients. 2021;13(4).PMID:33808209
17. Publich Health England (PHE). Composition of foods integrated dataset (CoFID). Available at: https://www.gov.uk/government/publications/composition-of-foods-integrated-dataset-cofid (accessed day/month/year). 2021
18. Pinchen H, Zhang L, Roe M, Church S, Traka M, Finglas P. UK Composition of Foods Labelling dataset. Available at https://fnnbri.quadram.ac.uk/labelling/ (accessed day/month/year) 2021
19. Simpson E, Bradley J, Poliakov I, Jackson D, Olivier P, Adamson AJ, et al. Iterative Development of an Online Dietary Recall Tool: INTAKE24. Nutrients. 2017;9(2).PMID:28208763
20. Bradley J, Simpson E, Poliakov I, Matthews JN, Olivier P, Adamson AJ, et al. Comparison of INTAKE24 (an Online 24-h Dietary Recall Tool) with Interviewer-Led 24-h Recall in 11-24 Year-Old. Nutrients. 2016;8(6).PMID:27294952
21. Foster E, Lee C, Imamura F, Hollidge SE, Westgate KL, Venables MC, et al. Validity and reliability of an online self-report 24-h dietary recall method (Intake24): a doubly labelled water study and repeated-measures analysis. J Nutr Sci. 2019;8:e29.PMID:31501691
22. Rowland MK, Adamson AJ, Poliakov I, Bradley J, Simpson E, Olivier P, et al. Field Testing of the Use of Intake24-An Online 24-Hour Dietary Recall System. Nutrients. 2018;10(11).PMID:30404170
23. Anderson JW, Baird P, Davis RH, Jr., Ferreri S, Knudtson M, Koraym A, et al. Health benefits of dietary fiber. Nutr Rev. 2009;67(4):188-205.PMID:19335713
24. Dhaka V, Gulia N, Ahlawat KS, Khatkar BS. Trans fats-sources, health risks and alternative approach - A review. J Food Sci Technol. 2011;48(5):534-41.PMID:23572785
25. Huang Y, Chen Z, Chen B, Li J, Yuan X, Li J, et al. Dietary sugar consumption and health: umbrella review. BMJ. 2023;381:e071609.PMID:37019448
26. Trafimow D. The attenuation of correlation coefficients: a statistical literacy issue. Teaching statistics. 2015;38(1):25-8
27. Lombard MJ, Steyn NP, Charlton KE, Senekal M. Application and interpretation of multiple statistical tests to evaluate validity of dietary intake assessment methods. Nutr J. 2015;14:40.PMID:25897837
28. Tomova GD, Arnold KF, Gilthorpe MS, Tennant PWG. Adjustment for energy intake in nutritional research: a causal inference perspective. Am J Clin Nutr. 2022;115(1):189-98.PMID:34313676
29. Team R. RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL http://www.rstudio.com/. 2020
30. Gamer M, Lemon J. irr: Various Coefficients of Interrater Reliability and Agreement_. R package version 0.84.1, <https://CRAN.R-project.org/package=irr>. IFPS; 2019.

31. Allaire J, Gandrud C, Russell K, Yetman C. networkD3: D3 JavaScript Network Graphs from R_. R package version 0.4, <https://CRAN.R-project.org/package=networkD3>. 2017
32. Fox J, Weisberg S. An R Companion to Applied Regression_, Third edition. Sage, Thousand Oaks CA. <https://socialsciences.mcmaster.ca/jfox/Books/Companion/>. 2019
33. Douglas B, Martin M, Ben B, Steve W. Fitting Linear Mixed-Effects Models Using
lme4. Journal of Statistical Software. 2015;67(1):1-48
34. Basso M. FRvalidation_against24HRs. Available: https://githubcom/BMelissa/FRvalidation_against24HRs. 2024
35. C Braga B, Nguyen PH, Aberman N-L, Doyle F, Folson G, Hoang N, et al. Exploring an Artificial Intelligence–Based, Gamified Phone App Prototype to Track and Improve Food Choices of Adolescent Girls in Vietnam: Acceptability, Usability, and Likeability Study. JMIR Formative Research. 2022;6(7):e35197.PMID:35862147
36. Svensson Å, Magnusson M, Larsson C. Overcoming Barriers: Adolescents' Experiences Using a Mobile Phone Dietary Assessment App. JMIR mHealth and uHealth. 2016;4(3):e92.PMID:27473462
37. Thornton L, Gardner LA, Osman B, Green O, Champion KE, Bryant Z, et al. A Multiple Health Behavior Change, Self-Monitoring Mobile App for Adolescents: Development and Usability Study of the Health4Life App. JMIR Formative Research. 2021;5(4):e25513.PMID:33843590
38. Sardi L, Idri A, Fernandez-Aleman JL. A systematic review of gamification in e-Health. J Biomed Inform. 2017;71:31-48.PMID:28536062
39. Amoutzopoulos B, Steer T, Roberts C, Collins D, Trigg K, Barratt R, et al. Rationalisation of the UK Nutrient Databank for Incorporation in a Web-Based Dietary Recall for Implementation in the UK National Diet and Nutrition Survey Rolling Programme. Nutrients. 2022;14(21).PMID:36364812
40. Wong L. Data analysis in qualitative research: a brief guide to using nvivo. Malays Fam Physician. 2008;3(1):14-20.PMID:25606106

Supplementary Files

Preprint JMIR Publications

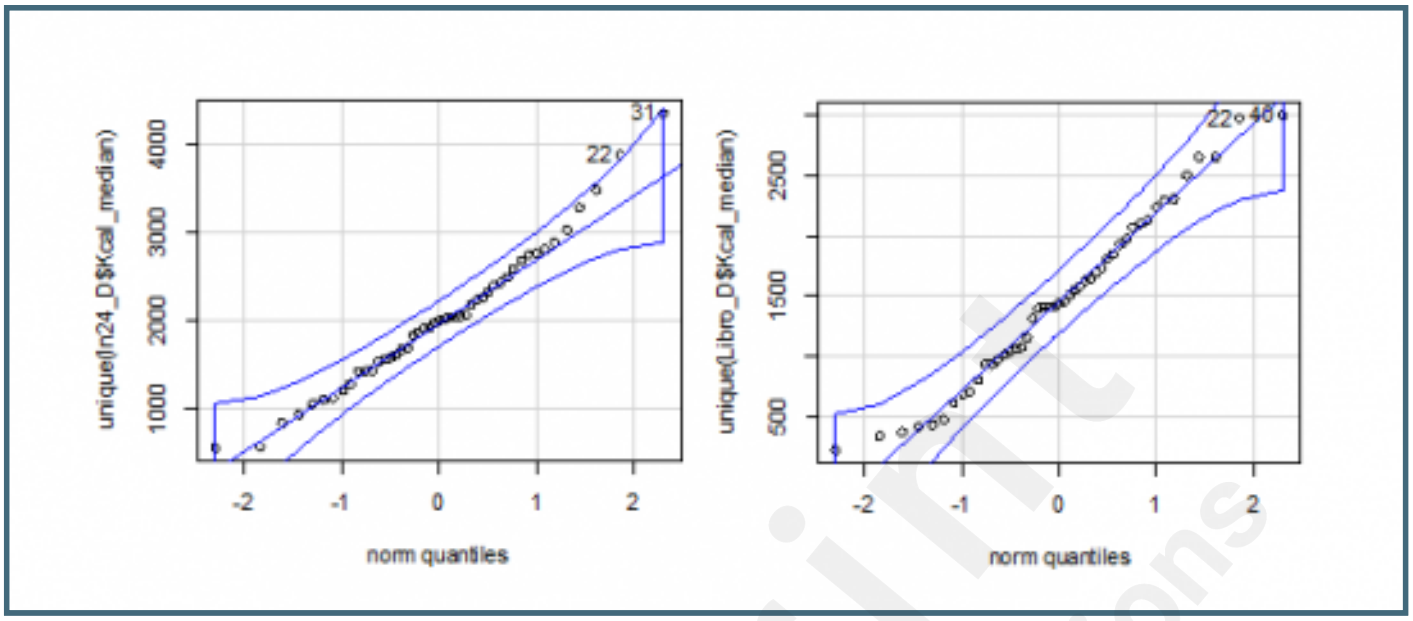
Figures

Preprint JMIR Publications

Customization of the FR program in the Libro app. Orange: the welcome message displayed before the recording week includes written and video instructions. These instructions remained accessible throughout the program. Green: User-interface (top) during a recording day, (bottom) during a non-recording day. Meals are displayed as a task which had to be ticked off for compliance checks via the Nutritics platform. Red: Example of the 4-5 push reminders during recording days (top) and 1during non-recording days (bottom).

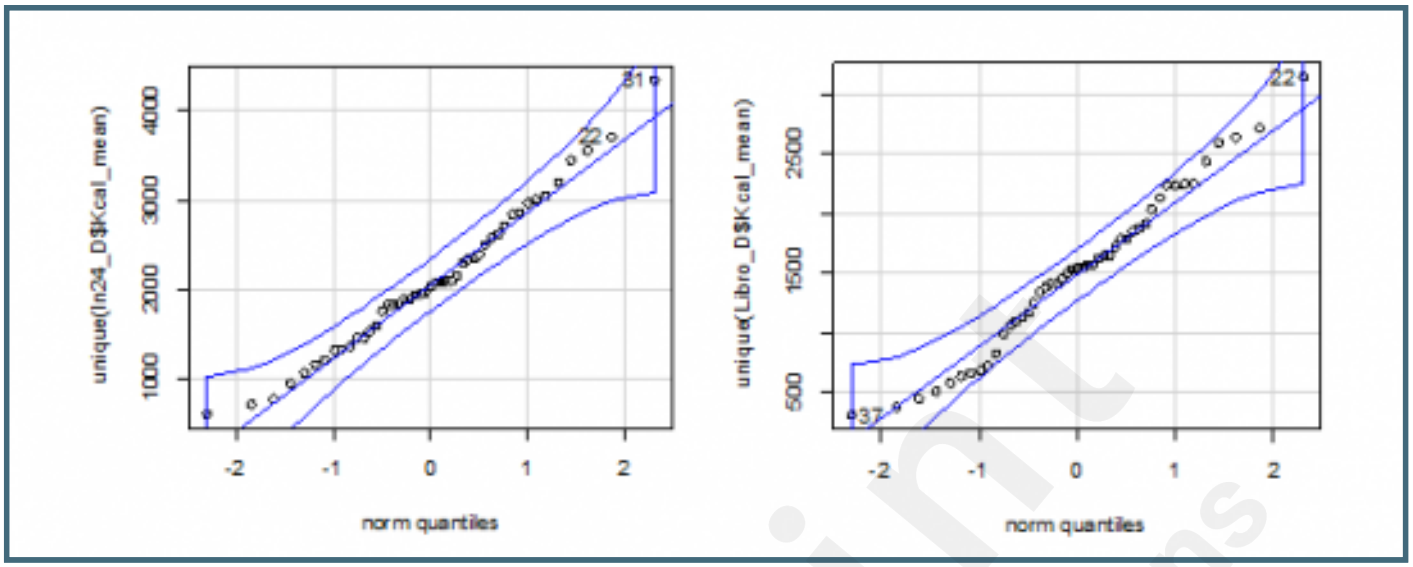
Example of number of layers and food recording process. Orange: layer 1; recording tasks are displayed, to start recording food participants are required to press the “go to log” button. Light orange: layer two; food log interface, to log food participants are required to press “+”. Yellow: layer 3; within each meal, participants can type food and select among options within a drop-down menu. Additional entry modalities were made possible (barcode, voice notes). Light yellow: layer 4; portion size selection, for some food items examples of portion size are shown. The same selection process is repeated per each food item consumed; all items are displayed in the grey window at the bottom and need to be saved before leaving the layer. Back to the layer 1, participants can tick the completed task.

Qq plots for median EI from single recalls for each participant (left: Intake24, right: Libro).



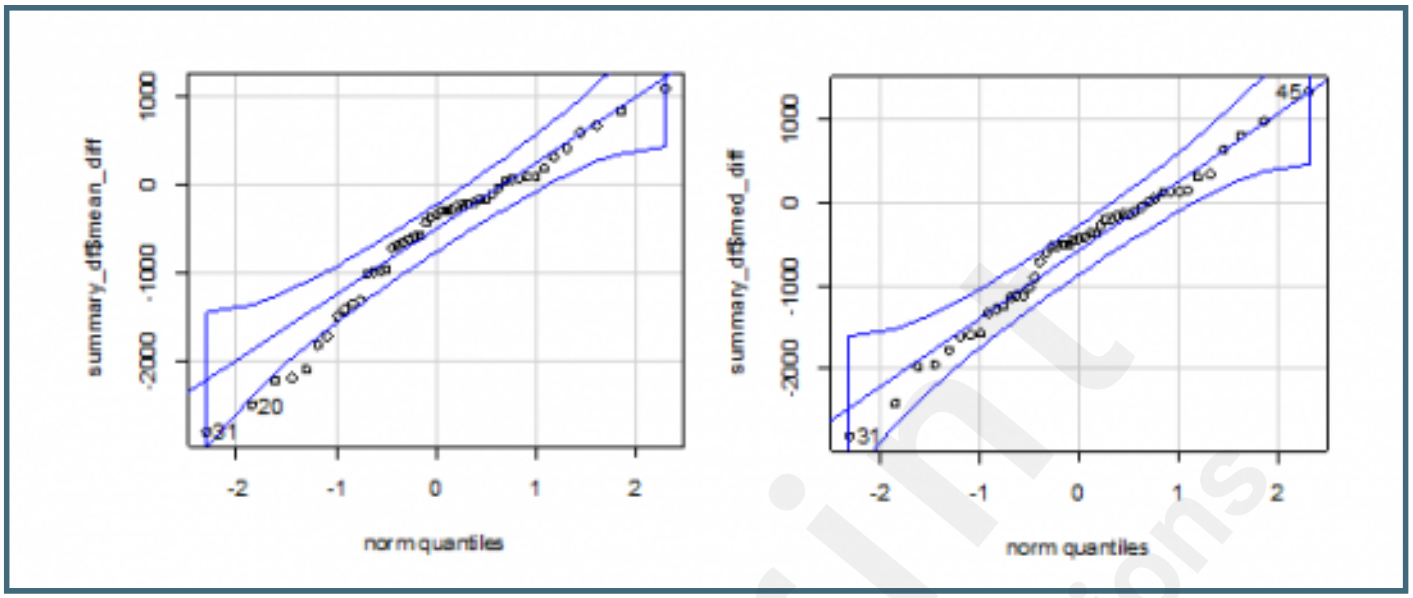
Preprint JMIR Publications

Qq plots for mean EI from single recalls for each participant (left: Intake24, right: Libro).



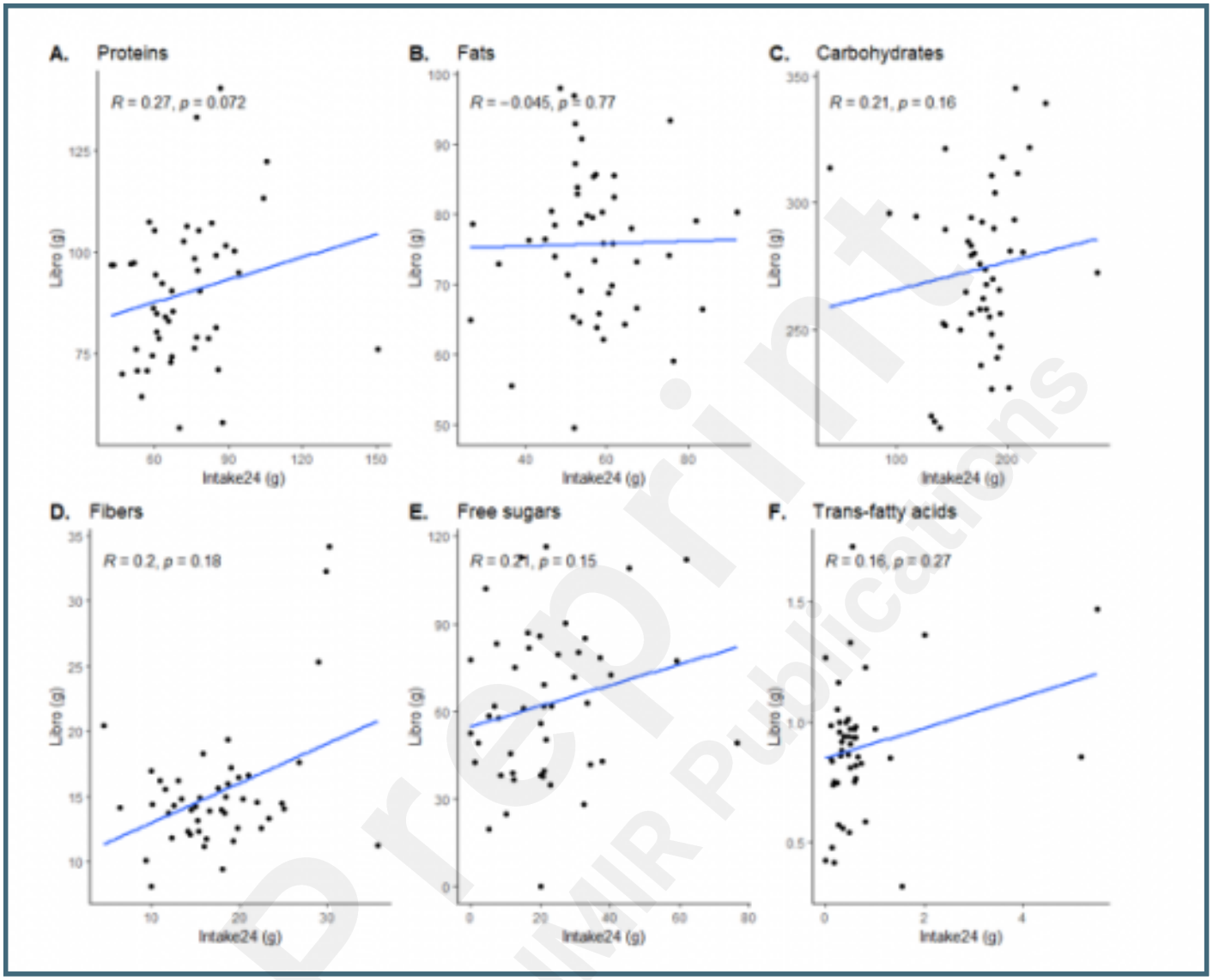
Preprint
JMIR Publications

Qq plots for differences between Libro and Intake24 intake using mean (left) and median (right) values of single recalls for each participant.



Preprint JMIR Publications

Scatter plot of Libro versus Intake24 mean values with a linear regression line for (A.) protein intake, (B.) fat intake, (C.) carbohydrates intake, (D.) fibre intake, (E.) free sugar intake, (F.) trans-fatty acids intake.



Preprint JMIR Publications

CONSORT (or other) checklists

Preprint JMIR Publications

GRRAS checklist for reporting of studies of reliability and agreement.
URL: http://asset.jmir.pub/assets/834566e9c0e31a20ad7c33c3ceb09784.pdf

Preprint JMIR Publications