nutrients

MDPI

Article

Performance of the digital dietary assessment tool MyFoodRepo

Claire Zuppinger [1,] *, Patrick Taffé [1], Gerrit Burger [1], Wafa Badran-Amstutz [1], Tapio Niemi [1], Clémence Cornuz [1], Fabiën N. Belle [1,2], Angeline Chatelan [2,3], Muriel Paclet Lafaille [4], Murielle Bochud [1] and Semira Gonseth Nusslé [1]

[1] Center for Primary Care and Public Health (Unisanté), University of Lausanne, Switzerland; patrick.taffe@unisante.ch (P.T.); gerrit.burger@student.uni-tuebingen.de (G.B.); wafa.badran-amstutz@unisante.ch (W.B.-A.); tapio.niemi@unisante.ch (T.N.); clemence.cornuz@vd.ch (C.C); murielle.bochud@unisante.ch (M.B.); semira.gonseth-nussle@unisante.ch (S.G.N.)

[2] Institute of Social and Preventive Medicine (ISPM), University of Bern, 3012 Bern, Switzerland; fabien.belle@ispm.unibe.ch (F.B.)

[3] Department of Nutrition and Dietetics, School of Health Sciences (HEdS-GE), University of Applied Sciences and Arts Western Switzerland (HES-SO); angeline.chatelan@hesge.ch (A.C.)

[4] Lausanne University Hospital (CHUV), Switzerland; muriel.lafaille@chuv.ch (M.L);

* Correspondence: claire.zuppinger@gmail.com (C.Z.); Tel.: +41-21-314-60-60

Received: date; Accepted: date; Published: date

Abstract: Digital dietary assessment devices could help overcome limitations of traditional tools to assess dietary intake in clinical and/or epidemiological studies. We evaluated the accuracy of the automated dietary app MyFoodRepo (MFR) against controlled reference values from weighted food diaries (WFD). MFR capabilities to identify, classify and analyze the content of 189 different records were assessed using Cohen and uniform kappa coefficients and linear regressions. MFR identified 98.0% ±1.5 of all edible components and was not affected by increasing numbers of ingredients. Linear regression analysis showed wide limits of agreements between MFR and WFD methods to estimate energy, carbohydrates, fat, proteins, fiber and alcohol contents of all records and a constant overestimation of proteins, likely reflecting the overestimation of portion sizes for meat, fish and seafood. MFR mean portion size error was 9.2% ±48.1 with individual errors ranging between -88.5% and +242.5% compared to true values. Beverages were impacted by the app's difficulties to correctly identify the nature of liquids (41.9% ±17.7 of composed beverages correctly classified). Compensation of under- and overestimations of weight by MFR, along with its strong segmentation and classification capabilities resulted in a generally good agreement between MFR and WFD, which would be suited for the identification of dietary patterns, eating habits and regime types.

Keywords: dietary assessment; accuracy; validation; food intake; diet; mobile food record; app

1. Introduction

Although diet is recognized as a large contributor to the onset and etiology of non-communicable diseases, its valid and reliable measurement in clinical and epidemiological studies remains a challenge, mainly because of its reliance on self-reported information and the lack of accessible tools to collect good quality information. Conventional dietary assessment tools are either of good scientific quality but involve high implementation costs (24-h recall) and substantial commitment from the participants (dietary records), or are easily implemented but lack accuracy and precision (food frequency questionnaires) [1].

Digital measurement devices can help overcome limitations of conventional dietary assessment tools and provide a cost-effective way to simplify and scale up nutritional data collection. Such devices were shown to increase user acceptance, while providing valuable real-time food intake data [2-4], and have the potential to eliminate participant burden linked to portion size estimation [5]. Digital image capture of foods is further facilitated by the distribution of mobile phones and population's familiarization with this technology. In 2017, more than 325'000 health mobile applications were available via the main app stores all over the world [6]. The majority of available diet mobile apps are often designed to support behavioral changes and either have not been validated for use in research or lack accuracy [7-10]. Although digital dietary assessment tools employed in research overcome these limitations, they often rely on more cumbersome participation from users, for example necessitating specific experimental settings [11], requiring participants to wear impractical gear (e.g. chest-worn camera) [12-14] or manually select dishes and estimate portion sizes [2, 15-17]. The accuracy of these tools relies on the comprehensiveness and quality of their underlying nutrition databases, whose continuous update is a challenge [3, 10]. The relevance of emerging digital dietary assessment tools for epidemiologic research is also made difficult by the scarcity of information relative to the tools' development process, the large variation in intake calculations and the differences of methodologies employed amongst validation studies [10]. Validation studies rarely assess all the different stages of dietary recognition: (1) segmentation (i.e. the ability of a tool to recognize <u>where</u> the different edible components of an image find themselves); (2) classification (i.e. the ability to correctly identify <u>what</u> the content of each segment is); (3) portion size estimation, as well as (4) energy and macronutrient calculation. Whereas some researchers investigating digital tools mainly focus on the segmentation and classification of food components – as is the case for DietCam [18] – other validation studies – e.g. for e-Ca [19] or mFR [20] – concentrate on determining weight error and related energy and macronutrient intakes. In most validation studies – amongst which the performance analysis of the digital dietary assessment tools Keenoa [9], PIQNIQ [2], EaT [21] and Bridge2U [22] – the accuracy of energy and macronutrient content is the sole endpoint investigated. One exception is Snap-n-Eat, a mobile phone recognition system able to perform automatic segmentation and classification of foods and allowing subsequent weight estimation and energy and macronutrient content calculation, but this app has not yet been validated [23].

The aforementioned validated tools often present low to moderate levels of accuracy [4, 24], showing wide limits of agreements compared to established dietary methods [9, 21, 22]. Newly developed digital dietary assessment tools need to be compared against a reference method (e.g. weighted food diaries over several days) and assessed among participants with similar dietary habits as the population of interest, which is not systematically the case.

In this context, we aimed to assess the accuracy of the automated dietary assessment device MyFoodRepo, by investigating its capabilities to identify, classify, estimate portion sizes and determine macronutrient content of diet against reference values from weighted food diaries.

2. Materials and Methods

MyFoodRepo (MFR) is a mobile application developed by the team of Prof. Marcel Salathé (Digital Epidemiology Lab - École Polytechnique Fédérale de Lausanne), which can be used to track food consumption from pictures of meals and beverages or from scanned food products' barcodes [25]. The app does not require any fiduciary marker for image recognition and its algorithm, based on thousands of images (May, 2021), uses artificial intelligence for image content analysis. The system incorporates an annotation interface, which allows textual conversation between MFR app users and a human reviewer from MFR app developers' team.

Three researchers conducted the present validation study, using the photography and barcode scanning features of MFR mobile app to record foods and beverages. MFR was evaluated on four different aspects: (1) segmentation: the ability to accurately differentiate the distinct edible components of a record (e.g. discerning the presence of a beige-colored segment from a red-colored segment in a plate while leaving aside the background and cutlery); (2) classification: the ability to correctly identify the content of each detected segment (e.g. identifying that the beige-colored segment is pasta and that the red-colored segment is Bolognese sauce); (3) portion size estimation: the accuracy of weight estimates for each detected and exactly classified segment; (4) overall performance: the accuracy and agreement of energy and macronutrient estimates compared to weighted food diaries.

2.1 Data collection

Data collection extended from September to December 2019. We aimed at gathering a minimum of 180 records – 1 record defined as either 1 photograph or 1 scan entered into MFR - and distributed as such: 60 *composite foods*, made up of ≥3 segments; 60 *simple foods*, made up of 1-2 segment(s); 30 *composite beverages*, made up of ≥2 segments; 30 *simple beverages*, made up of 1 segment only. For foods, a *segment* may refer to a single ingredient (e.g. carrots) or mixed ingredients that form a unified item to be recognized by MFR (e.g. ratatouille). For beverages, a *segment* refers to a single ingredient (e.g. tea). Records were arbitrarily selected by the researchers and did not represent daily intake. Industrial processed foods with a barcode were directly scanned into the app.

2.1.1 Controlled values measured from the weighted food diaries

To produce controlled values, we created, tested and optimized food diaries with the advice of a registered dietician. We entered each record into the food diaries. Weight and nutritional values from barcoded products were directly transcribed from their respective packaging and nutrition labels. Ingredients and complete segments from photographed records were carefully weighted and described. For cooked or mixed items, we noted the precise recipe.

Data from the weighted food diaries were analyzed by the dietician using the software PRODI 6.5 Swiss (Nutri-Science GmbH, Germany) and food composition databases, resulting in nutritional values being retrieved from the Swiss Food Composition Database [26], the French Food Composition Database [27] and the German Nutrient Database [28] to obtain energy and macronutrient content data on all records.

We additionally classified all segments into 37 *food types* (Table S1). Segments made up of mixed ingredients were classified as per the ingredient bringing the highest calorific content (e.g. potato gratin classified into "Tubers"). Food types were further coded into 23 *food groups* and again into 7 *food categories*, corresponding to categories of the Swiss food pyramid [29].

2.1.2 Measurements made by MFR

For each record entered into the weighted food diaries, a new picture or scan was saved into MFR. Records were processed by the MFR algorithm, and curated by the MFR app developers, who were able to ask clarifications about the entered records via the built-in annotation interface of the app. To test MFR's sole ability to recognize and analyze food content, the researchers were instructed not to leave any spontaneous descriptions about the content of the pictured foods in the app's annotation interface. However, researchers were allowed to answer ulterior questions if the MFR app developers asked for clarifications (e.g. "Did you put sugar in the tea?"; "Is it beef or veal on the picture?"). All MFR app developers performed their tasks while being blinded to the study, since they did not know if they had interactions with researchers from this validation study or participants from other ongoing studies.

MFR draws nutritional values of food and beverages from the Swiss Food Composition Database [26] and the French Food Composition Table, Ciqual [27]. Nutritional values of barcode scanned records are extracted from MFR's community-driven associated database, Open Food Repo [30], whose members can add and correct information from nutrition labels.

The researchers obtained the data extracted from MFR by the app developers' team on February 17, 2021, with details on date and time of collection, name, weight and/or volume, energy and macronutrient content of each detected segment contained in the records. The researchers consequently listed all segments identified by MFR app, and classified them into *food types*, *groups* and *categories*.

2.2 Data analysis

2.2.1 Segmentation

Segments (i.e., different components of a record) correctly identified by MFR were coded as *found* segments (F); overlooked segments were coded as *omissions* (O); additional segments erroneously identified by MFR and not actually present on a record were coded as *intrusions* (I). The segmentation percentage of accuracy was then calculated by the number of *found*, *omitted* and *intruded* segments respectively, over the total number of original segments in the weighted food diaries.

2.2.2 Classification

MFR naming of each *found* segment was compared to the corresponding true segment designation. MFR's classification performance was assessed by describing each match as *exact match*, *close match*, *far match* or *mismatch* according to the following criteria: *exact match* (E): MFR segment belongs to the correct *food type* and/or is labeled after a product meaning the exact same thing (e.g. cherry tomatoes vs. tomatoes); *close match* (C): MFR segment belongs to the correct *food type* but its naming is either too generic, not specific enough, refers to a slightly different product (e.g. beef vs. veal) or contains an overlooked ingredient within (e.g. tea with added sugar vs. tea); *far match* (F): MFR segment belongs to the wrong *food type* but to the correct *food group* (e.g. pasta vs. rice); *mismatch* (M): MFR segment belongs to the wrong *food type* and the wrong *food group* (e.g. carrot vs. potato).

Classification accuracy was calculated as the percentage of *exact*, *close*, *far* and *mismatches* among the total number of found segments. Proportion differences between MFR and controlled values were tested using Fisher's exact test.

To evaluate MFR categorization performance at different levels of granularity, it was assessed by main *food categories*, *food groups* and *food types* using the Cohen kappa as a reliability indicator, the uniform kappa coefficient as an agreement indicator, specificity and sensitivity calculation.

As inter-rater agreement between two researchers judging MFR segmentation and classification was high for 30 random records (uniform kappa Ku = 1 [95% confidence interval: 1;1] and Ku = 0.744 [0.607;0.843] respectively [31]), one researcher could proceed with coding the remaining records.

2.2.3 Portion size estimation

Weight error (difference between MFR weight estimation and true weight) was determined for each exactly classified segment. Mean weight, mean error, and mean absolute error were then calculated per *food type*, *food group* and *food category* and mean errors plotted into boxplots. Mean differences between true and estimates values were assessed with paired t-tests. Two-sided p-values ≤ 0.05 were considered significant. Finally, accuracy of portion size estimates was assessed dividing the mean estimated weight by the mean true weight for each *food type*, *group* or *category*.

2.2.4 Overall performance for energy and macronutrient content

To assess the agreement between the two methods, linear regression analysis was performed for energy and macronutrient estimates (fat, carbohydrates, proteins, fiber, alcohol). Linear regression was preferred (with MFR measurement as the dependent variable and weighted food diaries' measurements as the independent variable) to the commonly used Bland-Altman method, as the latter was shown to provide biased results when one of the two measurement methods has negligible measurement error [32]. Therefore, under the assumption that weighted food diaries correspond to an unbiased gold standard with negligible measurement error, we estimated the differential and proportional bias from the app, by regressing MFR measurements as a function of the controlled values. The 95% limits of agreement were then calculated by modeling the measurements' heteroscedasticity from the app [33].

To allow for comparison and give a general idea of the MFR data dispersion, we additionally calculated the coefficient of variation of MFR at different controlled values' (25th percentile, median, 75th percentile) as well as the mean coefficient of variations ($\overline{Cv}$) for energy and macronutrients.

3. Results

A total of 189 records were collected (63 *composite foods*, 63 *simple foods*, 30 *composite beverages*, 33 *simple beverages*). Among all records, 174 (92%) were recorded by photography, while 15 (8%) were barcode scanned. For practical reasons, only *simple foods* and *simple beverages* benefited from the scan feature in this study. Clarifications were demanded by the MFR app developers via the annotation interface in 43% (n=81) of all cases but exclusively for dishes and beverages recorded by photography (Figure S1).

Four probable weight transcription errors resulting in unrealistic entries in the food diaries were considered as outliers and removed from the portion size and overall performance analyses.

3.1. Segmentation

From 352 total true segments, MFR found 98.0%, omitted 2.0%, and intruded 1.4% (Table 1). *Simple beverages* had a 100% segmentation success rate and *composite beverages* only one omission with the oversight of the decorative candy on a cocktail. Among food records, the app omitted 6 segments: capers, horseradish sauce, pasta sauce, vinegar, dried tomatoes, and red peppers in vegetable mix respectively. The 5 intrusions, all from foods, corresponded to the erroneous addition of salad dressing on salads.

Table 1. Segmentation accuracy: proportion ± standard deviation (and number of segments) found, omitted, and intruded by MFR, by record type

Segments	Total (n=352)	Composite foods (n=208)	Simple foods (n = 79)	Composite beverages (n = 32)	Simple beverages (n = 33)
%Found (n)	98.0% ± 1.5 (n=345)	97.6% ± 2.1 (n=203)	98.7% ± 2.5 (n=78)	96.9% ± 6.1 (n=31)	100.0% ± 0.0 (n=33)
%Omitted (n)	2.0% ± 1.5 (n=7)	2.4% ± 2.1 (n=5)	1.3% ± 2.5 (n=1)	3.1% ± 6.1 (n=1)	0.0% (n=0)
%Intruded (n)	1.4% ± 1.2 (n=5)	1.5% ± 1.7 (n=3)	2.5% ± 3.5 (n=2)	0.0% (n=0)	0.0% (n=0)

Differences between *composite* and *simple* records were not significant, suggesting that MFR segmentation performance did not decrease with increasing records' complexity and number of ingredients.

3.2. Classification

Among all 345 segments found by MFR, 87.5% ±3.5 were classified as exact match, 8.4% ±3.0 as close match, 1.2% ±1.1 as far match and 2.9% ±1.8 as mismatch (Table 2).

Scanned records showed perfect classification accuracy, with 100% of segments classified as exact match.

Best results were found for records where clarifications had been demanded by the MFR app developers via the built-in annotation interface, compared to records where no clarification had been asked, with a slightly higher percentage of exact and close matches (98.4% ± 1.8 vs. 93.0% ± 4.0) and a lower percentage of far matches and mismatches (1.6% ± 1.8 vs. 7.0% ± 4.0) ($p = 0.01339$).

Simple records showed strong results with only 2 mismatches within *simple beverages* and 3 close matches within *simple foods*, the remaining *simple* records being classified as exact match. *Composite* records presented more mitigated outcomes. MFR performed the poorest among *composite beverages* with only 41.9% (± 17.7) exact matches. Close matches within *composite beverages* were mainly due to the omission of milk and/or sugar in hot beverages and the description of cocktails, judged too generic, whereas mismatches resulted from the erroneous classification of alcoholic beverages as soft drinks and soft drinks as non-alcoholic non-sweetened beverages.

Table 2. Classification accuracy: proportion ± standard deviation (and number of found segments) classified as exact match, close match, far match, and mismatch, by record type

Records' classification	Total (n=345)	Composite foods (n=203)	Simple foods (n = 78)	Composite beverages (n = 31)	Simple beverages (n = 33)
%Exact match (n)	87.5% ± 3.5 (n=302)	90.1% ± 4.1 (n=183)	96.2% ± 4.3 (n=75)	41.9% ± 17.7 (n=13)	93.9%± 8.3 (n=31)
%Close match (n)	8.4% ± 3.0 (n=29)	6.9% ± 3.5 (n=14)	3.8% ± 4.3 (n=3)	38.7% ± 17.4 (n=12)	0.0% (n=0)
%Far match (n)	1.2% ± 1.1 (n=4)	2.0% ± 1.9 (n=4)	0.0% (n=0)	0.0% (n=0)	0.0% (n=0)
%Mismatch (n)	2.9% ± 1.8 (n=10)	1.0% ± 1.4 (n=2)	0.0% (n=0)	19.4%± 14.1 (n=6)	6.1% ± 8.3 (n=2)

This influenced MFR classification results among *food groups* (**Table 3**), with the lowest sensitivity percentages observed among "non-alcoholic sweetened beverages" (75%) and "alcoholic beverages" (76.2%). Apart from small reductions in the *food groups* "sweeteners" (83.3%), "fats & oils" (85.7%) and "potatoes, legumes and beans" (90.5%), sensitivity percentages were otherwise high.

Table 3. Classification accuracy: Cohen Kappa, uniform Kappa, sensitivity and specificity of MFR classification compared to controlled values from weighted food diaries, by food groups

Food Groups	Cohen Kappa		Uniform Kappa		Sensitivity [%]		Specificity [%]	
	Kappa	Std. Err.	Kappa	[95% CI]	Sensitivity	[95% CI]	Specificity	[95% CI]
Non-alcoholic non-sweetened beverages	0.8607	0.0533	0.977	[0.954;0.994]	100	[75.3;100]	98.8	[96.9; 99.7]
Vegetables	1.0000	0.0538	1	[1;1]	100	[95.3;100]	100	[98.6;100]
Fruit	1.0000	0.0538	1	[1;1]	100	[85.8;100]	100	[98.9;100]
Juice	0.7721	0.0524	0.977	[0.954;0.994]	100	[59;100]	98.8	[97.0;99.7]
Meat & poultry	1.0000	0.0538	1	[1;1]	100	[83.9;100]	100	[98.9;100]
Fish & seafood	1.0000	0.0538	1	[1;1]	100	[54.1;100]	100	[98.9;100]
Unclassified meat	1.0000	0.0538	1	[1;1]	100	[39.8;100]	100	[98.9;100]
Eggs, plant-based protein products	0.8874	0.0535	0.994	[0.983;1]	100	[39.8;100]	99.7	[98.4;100]
Dairy products (excl. milk)	1.0000	0.0538	1	[1;1]	100	[80.5;100]	100	[98.9;100]
Milk & milk-based beverages	1.0000	0.0538	1	[1;1]	100	[29.2;100]	100	[98.9,100]
Seeds & nuts	1.0000	0.0538	1	[1;1]	100	[29.2;100]	100	[98.9;100]
Fats & oils	0.8542	0.0538	0.988	[0.971;1]	85.7	[42.1;99.6]	99.7	[98.4;100]
Cereals and cereal-based products	0.9598	0.0538	0.988	[0.971;1]	96.3	[81; 99.9]	99.7	[98.3;100]
Rice, rice-based products	0.9319	0.0537	0.994	[0.983;1]	100	[59;100]	99.7	[98.4;100]
Potatoes, legumes & beans	0.9469	0.0538	0.988	[0.971;1]	90.5	[69.6;98.8]	100	[98.9;100]
Salty snacks	1.0000	0.0538	1	[1;1]	100	[54.1;100]	100	[98.9;100]
Sweet dishes	1.0000	0.0538	1	[1;1]	100	[78.2;100]	100	[98.9;100]
Sweeteners	0.9076	0.0536	0.994	[0.983;1]	83.3	[35.9;99.6]	100	[98.9;100]
Non-alcoholic sweetened beverages	0.8122	0.0536	0.977	[0.954;0.994]	75.0	[42.8;94.5]	99.7	[98.3;100]
Alcoholic beverages	0.8574	0.0533	0.971	[0.936;0.994]	76.2	[52.8;91.8]	100	[98.9;100]
Condiments & sauces	0.9665	0.0538	0.988	[0.971;1]	97.0	[84.2;99.9]	99.7	[98.2;100]
Milk substitutes	1.0000	0.0538	1	[1;1]	100	[2.5;100]	100	[98.9;100]
Soups	1.0000	0.0538	1	[1;1]	100	[69.2;100]	100	[98.9;100]

Uniform kappa results were high among *food groups*, laying between 0.971 [0.936;0.994] and 1 [1;1] and showing a strong classification agreement between MFR and controlled values. Cohen

Kappa results, which measure reliability between the 2 methods, were prone to more fluctuation due to their high dependence to the number of segments in each discriminated *food group*.

Globally, classification reliability and agreement of MFR compared to controlled values were nevertheless high in all three levels of classification granularity: *food categories (0.963)*, *food groups* (0.9554) and *food types* (0.9559) (Table 4).

Table 4. Global classification reliability (Cohen kappa) and agreement (uniform kappa) between MyFoodRepo and controlled values from weighted food diaries

Level of granularity	Cohen Kappa		Uniform Kappa	
	Kappa	**Std. Err.**	**Kappa**	**[95% CI]**
Food Categories	0.9603	0.0254	0.963	[0.943; 0.983]
Food Groups	0.9554	0.0158	0.958	[0.933; 0.979]
Food Types	0.9559	0.0145	0.958	[0.934; 0.979]

Cohen Kappa, uniform Kappa, sensitivity and specificity by *food types* and *food categories* can be found in the Supplementary material (Table S2 and Table S3).

3.3 Portion size estimation

Mean true weight of all exactly classified segments (n = 302) was 116.8g ± 92.0 whereas mean estimated weight was 114.4g ± 83.0 (p = 0.424), with a mean error of -2.4g ± 51.8. Nevertheless, mean absolute error was 32.8% and range of percentage error fluctuated between -88.5% and 242.5% of true weight, indicating that underestimations and overestimations by MFR compensate one another.

MFR portion size estimation performance by *food group* is shown in

nutrients

MDPI

Table 5.

nutrients

MDPI

Table 5. MFR portion size estimation performance: mean estimated weights vs. true weights, mean error, and mean absolute error of all exactly classified segments, displayed by food groups (first half)

Food Groups	Mean weight [g ± sd]		Ratio of MFR estimate to true value [1]	p-value [2]	Number of values n	Mean error [3] [g ± sd]	Mean absolute error [4] [g ± sd]	Mean absolute error [4] [%]	Range of error percentage [5] [min% ; max%]
	True value	MFR estimate							
Non-alcoholic non-sweetened beverages	237.3 ± 88.2	192.9 ± 45.0	0.81	0.052	7	-44.4 ± 48.7	59.9 ± 22.3	35.8	-33.3 ; 117.4
Vegetables (incl. roots)	82.3 ± 45.3	76.1 ± 43.7	0.92	0.073	74	-6.2 ± 29.3	20.8 ± 21.4	29.2	-88.5 ; 150.0
Fruits	127.0 ± 70.0	120.4 ± 109.3	0.95	0.682	24	-6.6 ± 78.1	47.5 ± 61.6	37.1	-77.3 ; 117.4
Juice (non-sweetened)	181.8 ± 137.4	211.3 ± 134.7	1.16	0.122	8	29.5 ± 47.4	42.0 ± 35.0	32.4	-33.3 ; 66.7
Meat & Poultry	95.6 ± 33.8	115.6 ± 39.7	1.21	0.0001**	18	20.0 ± 16.9	20.9 ± 15.7	23.3	-9.1 ; 51.9
Fish & Seafood	114.7 ± 34.5	160.0 ± 52.2	1.29	0.004**	6	45.3 ± 21.8	45.3 ± 21.8	39.6	17.6 ; 60.0
Unclassified meat	149.3 ± 65.8	147.5 ± 37.7	0.99	0.973	4	-1.75 ± 96.6	68.3 ± 55.8	46.6	-54.7 ; 86.9
Eggs & meat substitutes	82.3 ± 24.2	96.3 ± 27.5	1.17	0.027*	4	14.0 ± 6.9	14.0 ± 6.9	17.3	7.8 ; 22.2
Dairy products (excl. milk)	113.1 ± 134.1	99.3 ± 86.3	0.88	0.412	15	-13.8 ± 63.2	28.9 ± 57.5	42.5	-46.1 ; 200.0
Milk & milk-based beverages	255.0 ± 8.7	233.3 ± 28.9	0.91	0.423	3	-21.7 ± 37.5	21.7 ± 37.5	8.2	-24.5 ; 0.0
Seeds & nuts	33.7 ± 36.7	25.7 ± 14.0	0.76	0.632	3	-8.0 ± 24.8	16.0 ± 18.0	45.0	-47.4 ; 78.6
Fats & oils	12.5 ± 6.9	11.0 ± 2.2	0.88	0.570	5	-1.5 ± 5.4	4.3 ± 3.0	63.7	-33.3 ; 233.3

sd = standard deviation
[1] Ratio of MFR weight estimations over true weights. Values > 1 indicate an overestimation; values < 1 indicate an underestimation.
[2] Two-sided paired t-test between mean true weight and mean MFR weight estimates.
[3] Error calculated as: (MFR weight estimation – True weight) for each exactly classified segment
[4] Absolute error calculated as | (MFR weight estimation – True weight) | for each exactly classified segment
[5] Minimum and maximum percentage of error observed in each food group.
*p-value ≤ 0.05
**p-value ≤ 0.005

Table 6. MFR portion size estimation performance: mean estimated weights vs. true weights, mean error, and mean absolute error of all exactly classified segments, displayed by food groups (second half)

Food Groups	Mean weight [g ± sd]		Ratio of MFR estimate to true value [1]	p-value [2]	Number of values n	Mean error [3] [g ± sd]	Mean absolute error [4] [g ± sd]	Mean absolute error [4] [%]	Range of error percentage [5] [min% ; max%]
	True value	MFR estimate							
Cereals & cereal-based products	192.6 ± 135.2	164.5 ± 82.7	0.85	0.163	23	-28.1 ± 93.3	60.2 ± 75.8	31.9	-60.5 ; 104.5
Rice, rice-based products	96.3 ± 45.9	89.7 ± 42.3	0.93	0.362	7	-6.6 ± 17.6	14.9 ± 10.1	18.1	-23.1 ; 33.3
Potatoes, legumes & beans	129.8 ± 51.4	156.6 ± 75.3	1.21	0.023*	19	26.8 ± 46.8	36.8 ± 39.0	30.4	-18.9 ; 100.0
Salty Snacks	34.2 ± 10.5	40.3 ± 9.0	1.18	0.184	6	6.1 ± 9.7	6.1 ± 9.7	24.4	0.0 ; 100.0
Sweet dishes	57.9 ± 47.4	55.7 ± 40.9	0.96	0.487	13	-2.3 ± 11.4	7.3 ± 8.8	30.9	-44.4 ; 140.0
Sweeteners	71.4 ± 87.4	94.0 ± 115.7	1.32	0.162	5	22.6 ± 29.5	24.6 ± 27.4	57.9	-50.0 ; 100.0
Non-alcoholic sweetened beverages	263.3 ± 111.8	234.4 ± 71.3	0.89	0.327	9	-28.9 ± 83.1	30.0 ± 82.7	7.3	-50.0 ; 2.6
Alcoholic beverages	117.1 ± 62.7	129.0 ± 82.9	1.10	0.275	10	11.9 ± 32.4	23.9 ± 23.9	21.5	-41.2 ; 60.0
Condiments and sauces	72.8 ± 55.5	73.5 ± 50.2	1.01	0.932	26	0.7 ± 40.6	32.1 ± 24.0	57.5	-66.7 ; 242.5
Milk substitutes [6]	250.0 ± NA	250.0 ± NA	1.00	NA	1	0.0 ± NA	0.0 ± NA	NA	0 ; 0.0
Soup	237.4 ± 78.8	213.3 ± 48.0	0.90	0.269	9	-24.0 ± 60.7	44.0 ± 46.3	19.2	-42.4 ; 60.0

sd = standard deviation

[1] Ratio of MFR weight estimations over true weights. Values > 1 indicate an overestimation; values < 1 indicate an underestimation.

[2] Two-sided paired t-test between mean true weight and mean MFR weight estimates.

[3] Error calculated as: (MFR weight estimation – True weight) for each exactly classified segment

[4] Absolute error calculated as | (MFR weight estimation – True weight) | for each exactly classified segment

[5] Minimum and maximum percentage of error observed in each food group.

[6] Only one observation in the "milk substitutes" food group

*p-value ≤ 0.05

**p-value ≤ 0.005

Among all 23 *food groups*, half presented a mean absolute error between 25% and 50%. "Milk and milk-based beverages", as well as "Non-alcoholic sweetened beverages" had a mean absolute error below 10%. On the other hand, "Fats & oils", "Sweeteners" and "Condiments & sauces" showed a mean absolute error over 50%.

MFR significantly overestimated weight for "Meat & Poultry" ($p = 0.0001$), "Fish & Seafood" ($p = 0.004$), "Eggs & meat substitutes" ($p = 0.027$) and "Potatoes, legumes and Beans" ($p = 0.023$) (

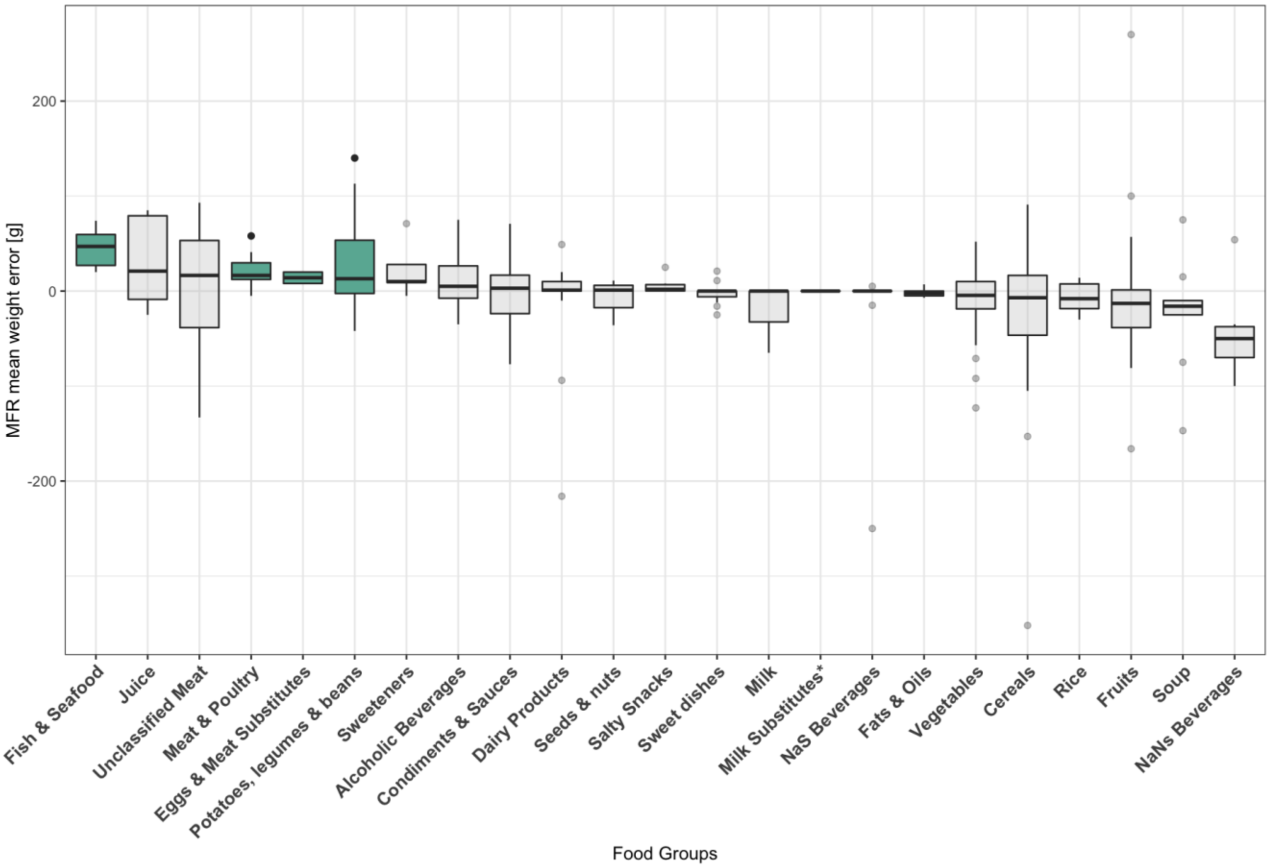

Figure **1**).

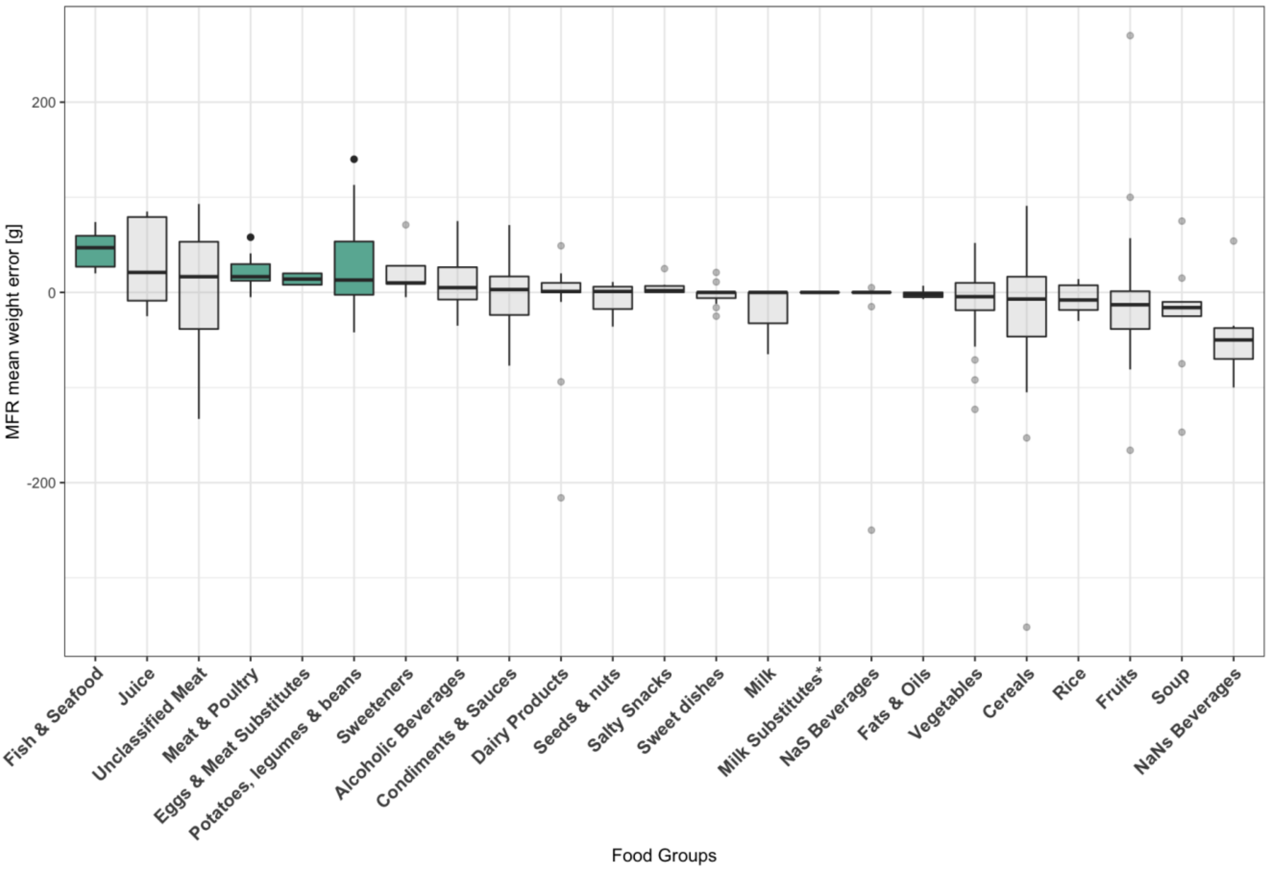

Figure 1. Mean weight errors per food group (NaNs: non-alcoholic non-sweetened; NaS: Non-alcoholic sweetened). Boxplots give median, interquartile range (IQR) and maximum 1.5 IQR. Colored boxplots indicate significant mean differences between estimated and true values (two-sided p-value ≤ 0.05). Four weight transcription errors resulting from unrealistic entries in the food diaries were removed from portion size analysis (not shown).

*Only one observation in the "milk substitutes" food group.

Comparable analyses calculated by *food types* and *food categories* can be accessed in the Supplementary Material (Table S4 and Table S5). No significant differences were observed between estimated and true mean weight at the food category level.

3.4 Overall performance for energy and macronutrient content

The overall performance analysis included all 185 records. Linear regression performed on MFR measurements versus controlled values from the food diaries show an overall overestimation by the app at small true values of energy and macronutrients and an underestimation tendency at higher true values (Figure 2). The y intercept was 113.3kcal for energy, 5.7g for fat, 20.4g for carbohydrates, 1.8g for fibers, and 0.6g for alcohol.

Only the linear regression line for protein content fell above the 1:1 line, indicating a systematic overestimation of proteins by MFR.

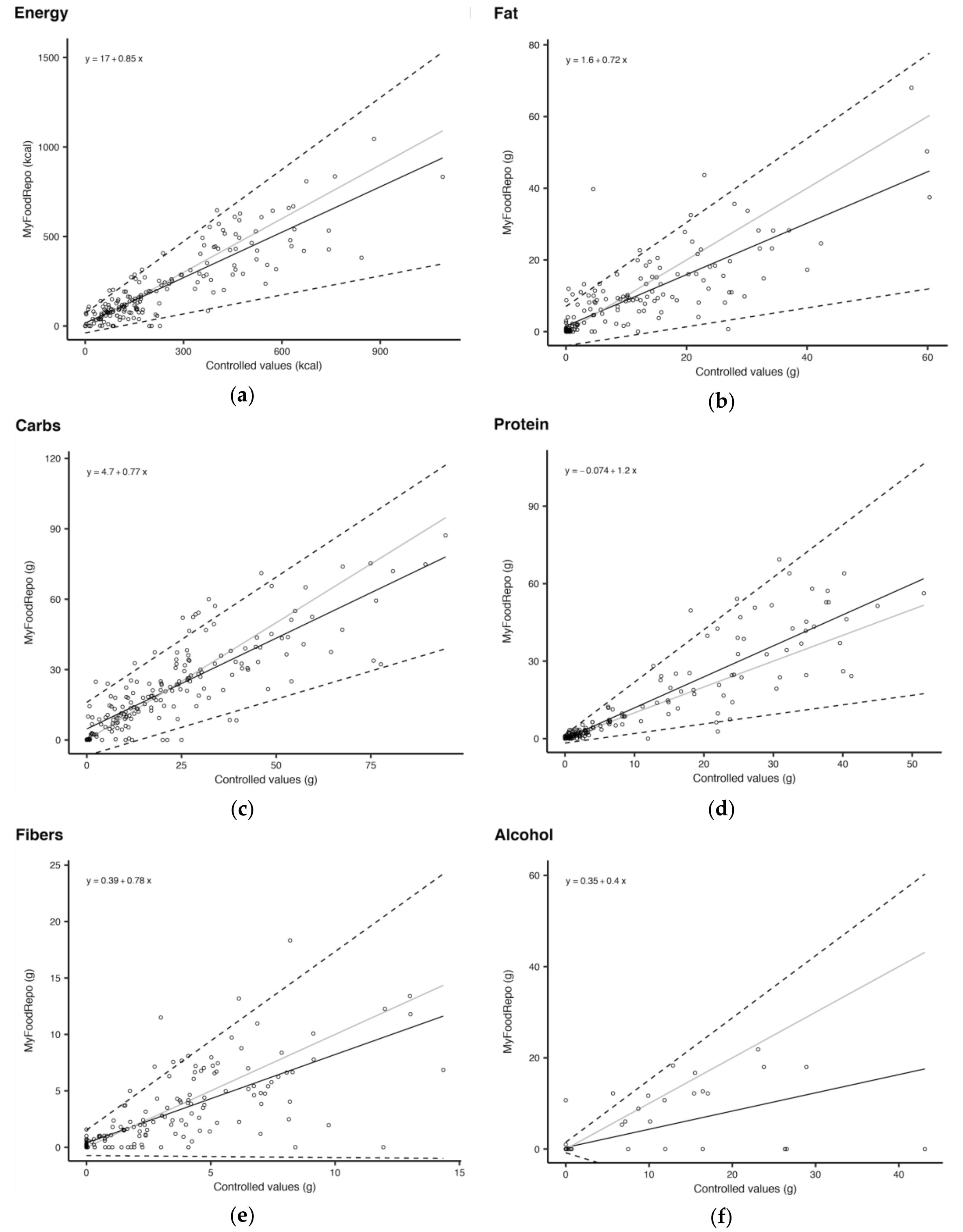

Figure 2. Overall performance for energy and macronutrient content: Linear regression of MFR versus controlled values for all found records, for content of (a) energy; (b) fat; (c) carbohydrates; (d) protein; (e) fiber and (f) alcohol. (Black line: linear regression line; dotted line: 95% limits of agreement; grey line: y=x).

Coefficients of variation (Cʋ) and mean coefficients of variation ($\overline{\text{Cʋ}}$) for energy and macronutrients of all records also suggest important levels of dispersion from MFR estimates (Table 7). As per the coefficients of variation (Cʋ) calculated at the 25th percentile, median and 75th percentile of true values, MFR's accuracy increased with increasing true energy and macronutrient content, meaning that the dispersion was higher for small quantities. Cʋ at the 25th percentile were particularly high for protein (1.96), alcohol (1.70), fat (1.68) and fibers (1.47). Fibers and alcohol had the highest $\overline{C\upsilon}$, whereas carbohydrates had the lowest $\overline{C\upsilon}$.

Table 7. Overall performance for energy and macronutrient content: Coefficient of variation and mean coefficient of variation of all MFR estimates calculated at the 25th percentile, median, and 75th percentile of controlled values for energy, fat, carbohydrates, fiber, and alcohol

	Coefficient of variation Cυ			**Mean coefficient of variation $\overline{C\upsilon}$**
	At true values' 25th percentile	**At true values' Median**	**At true values' 75th percentile**	**All records**
Energy	0.58	0.45	0.37	0.35
Fat	1.68	0.83	0.52	0.42
Carbohydrates	0.65	0.45	0.37	0.31
Protein	1.96	0.63	0.40	0.38
Fiber	1.47	0.72	0.62	0.58
Alcohol	1.70	1.70	1.70	1.25

Mean coefficients of variation were higher for beverages compared to food records, except for alcohol and to a lesser extent, carbohydrates (Table 8**Table 8**). The higher alcohol $\overline{C\upsilon}$ in foods comes from the unaccounted alcohol content in some sauce recipes, showing MFR' inability to identify sauces' composition. Beverages $\overline{C\upsilon}$ negatively affected $\overline{C\upsilon}$ of all records, especially for fat and fibers.

Table 8. Overall performance for energy and macronutrient content by record type: Mean coefficient of variation of MFR estimates for energy, fat, carbohydrates, fibers and alcohol by record type.

	Mean coefficient of variation $\overline{C\upsilon}$		
	All records	**Foods**	**Beverages**
Energy	0.35	0.33	0.75
Fat	0.42	0.41	1.18
Carbohydrates	0.31	0.32	0.27
Protein	0.38	0.37	0.65
Fibers	0.58	0.52	2.01
Alcohol	1.25	3.14	1.23

Linear regression figures for food and beverages separately can be found in the Supplementary Materials (Figure S2 and Figure S3).

4. Discussion

The purpose of this study was to assess the accuracy of the smartphone application MFR against weighted food diaries, which are currently considered the gold standard for dietary assessment. To our knowledge, this is the first study validating an automated digital dietary assessment tool by distinctly examining its different stages of food and beverage recognition, namely segmentation, classification, portion size estimation and energy and macronutrient content calculation.

Compared to most of its digital dietary assessment counterparts used in research, MFR requires minimal user input to record diet. While other digital dietary assessment tools require preliminary groundwork – for instance necessitating specific experimental settings with a fixed background [11] – others involve a fiduciary marker to be placed on the image [15-17] or for the users to delineate segments themselves in their specific tool [34, 35]. Despite only relying on a smartphone camera, MFR showed strong segmentation capacity, identifying 98.0% of all segments present. This did not only include visible items, as for similar technologies [23] but also blended or mixed segments. Additionally, segmentation accuracy did not significantly decrease for complex records, unlike observations made by the automated dietary tool, DietCam [5, 18].

MFR also performed generally well to classify found segments into *food types*, *food groups* and *food categories*. Scanned records showed perfect segmentation and classification results, bypassing the well reported reduction in accuracy and reliability of dietary assessment tools associated with non-exhaustive databases [3, 10]. Scanned items in MFR are indeed directly associated with the Open Food Repo database [30], which currently gathers more than 370'000 barcoded products sold in Switzerland. The database is open-access and user-enriched, ensuring a continuous update and alignment with population dietary habits. Uniform kappas over 0.958 indicated a good classification agreement between methods, at all levels of granularity. Percentages of exact matches exceeded 90% for *composite foods*, *simple foods*, and *simple beverages*, but only reached 41.9% *for composite beverages*, which could partially explain the large coefficients of dispersion for energy and macronutrients among beverages. "Alcoholic beverages", "non-alcoholic sweetened beverages" and "non-sweetened non-alcoholic beverages" were classified interchangeably and additions of milk and/or sugar in tea and coffee were often overlooked.

Unlike MFR, many digital dietary assessment tools used in studies rely on user participation for portion size estimation, either via a portion size selector [2, 19] or a complementary portion size booklet [22]. Portion size estimation relying on the capture of a single image was proven to reduce user burden as automated estimations are not affected by the user's lack of knowledge about quantities [36, 37]. Photography can also decrease data collection time and participant's disturbances in complex settings such as school cafeterias [11] as well as facilitate study implementation in environments with lower health and nutrition literacy or language barriers [38, 39]. MFR would nonetheless benefit from more precise estimations of portion sizes. Although the global mean portion size error of -2.4g ± 51.8 or 9.2% ± 48.1% was seemingly small, it resulted from the compensation of under- and overestimations of individual segments. With individual errors varying between -88.5% and +242.5% of true weight, the error range was wider than observed in the existing literature. In comparison, the electronic mobile-based food record e-Ca showed a mean error of 3% with errors ranging between -38% and +130% of true weights across 20 food and beverages displayed in a controlled setting [19], whereas mFR app, developed by Lee et al., found a minimum error of -38% and maximum error of 26% between automatically determined portion weights and control weights of 19 individual foods [20, 24]. Nonetheless, the relatively small number of participants and items assessed in the aforementioned studies reduce the likelihood of extreme errors compared to the present study.

MFR performance was particularly challenged by small or hidden ingredients within records. The greatest mean absolute weight errors were observed in the "Fats & oils", "Sweeteners" and "Condiments & sauces" *food groups*. Such imperceptible elements (e.g. sugar, oils and sauces) were also harder to classify by the app and showed weaker classification sensitivity compared to other *food groups*, an inevitable limitation of dietary data collection by photography [24]. The same conclusion can be extended to segmentation, where omissions and intrusions made by MFR mainly affected subsidiary food items, such as capers, sauces, or vinegar, as well as two additional segments blended in vegetable mixes.

The segmentation, classification and portion size estimation findings all influence the overall performance of MFR. We observed higher coefficients of variation $C\upsilon$ for energy and macronutrient estimates when true quantities were small, with a tendency towards overestimation. After a certain threshold, MFR underestimated all macronutrients with the exception of proteins. MFR overestimation tendency of proteins could be exacerbated by the significant weight overestimation of segments of "Meat & Poultry", "Fish & Seafood" as well as "Eggs and Meat substitutes". While carbohydrate estimates of all records showed reasonable results $(\overline{C\upsilon} = 0.31)$, fiber and alcohol had the highest mean coefficient of dispersion globally (0.58 and 1.25 respectively), especially in the case of beverages. Overall, linear regression analysis showed wide limits of agreements between MFR and weight record control method for energy, fat, carbohydrates, proteins, fiber and alcohol content of all records. Wide limits of agreement between a novel method and a control method are commonly observed in similar studies, whose digital dietary assessment tools are often validated for a utilization at the group level [9, 21, 22].

Nonetheless, our methodology focused on MFR's performance as a dietary assessment device, with no consideration regarding true daily dietary intake and real-life conditions (i.e., study participants taking pictures of their food on selected days). This made the comparison with other digital tools difficult and restricted our analysis to specific record's energy and macronutrient accuracy and precision. To avoid discriminating MFR for erroneously classifying or forgetting segments, we assessed weight errors on exactly classified segment only, which constitutes another limitation of our work. Furthermore, the decision not to use the app's comment fields during data collection may have reduced the accuracy of MFR. Indeed, MFR users are normally able to provide spontaneous description or comments in these integrated annotation fields, but we intentionally ignored this tool in the present study, in order to test MFR´s sole capability to identify and classify records' content.

5. Conclusion and recommendations

In light of the above, we would advise caution in the analysis of energy and macronutrient content for precise individual dietary assessment. Good agreement for portion size estimation between MFR and weight food diaries, along with the app strong segmentation and classification capabilities appears to be nonetheless suited for the identification of dietary patterns, eating habits, and regime types.

Statistical recalibrations to adjust for measurement error could potentially be used to improve MFR's current estimations. Energy adjustments could also be applied to increase the overall accuracy of MFR. This analytic method, which helps mitigating the effects of measurement errors when data are collected via a self-reported dietary assessment tool, has been assessed and applied in similar validation studies and could constitute the subject of subsequent research, provided that total energy intake is assessed [22, 40].

Currently, MFR's energy and macronutrient assessment is highly affected by imprecise portion size estimation. Improving portion size estimation capabilities would therefore prove valuable to strengthen the app's general performance. Combined with MFR's user-friendly recording interface, this would distinguish the app from other digital dietary assessment tools currently available for research purposes. Supported by a significant classification improvement with annotators' intervention, we would recommend MFR developers to focus on beverage content identification, to enhance MFR classification accuracy. The presence of alcohol, milk or sugar in beverages should be of particular focus and could be flagged by systematically asking participants for content of their beverages. This is for instance applied in the mobile device food record mpFR, which allows users to rectify misclassified segments before confirmation of intake [41, 42]. MFR already features an optional field for remarks visible during record entry. It would be in the app user's best interest to benefit from systematic prompts to ensure a more accurate classification of composite beverages. The same recommendation could be made for sauces and condiments.

The aforementioned adaptations could be put to the test in a subsequent study, further investigating MFR use in real-life settings with the measurement of daily dietary intake from study participants. In such conditions, and in order to fully compare MFR performance and practical implementation in epidemiological studies over traditional dietary assessment methods, researchers should assess the relevance of participants' notes, potential prompts or the use of a fiduciary marker on the pictures for portion size estimation and energy and macronutrient calculation. Tradeoffs in terms of time, cost-efficiency and practicability should nevertheless be considered, to avoid increasing user burden.

6. Patents

Supplementary Materials: The following are available online at www.mdpi.com/xxx/s1, Table S1 : Description of food types and corresponding food groups and food categories, Figure S1: Characteristics of records collected, Table S2: Cohen Kappa, uniform Kappa, sensitivity and specificity of MFR classification compared to controlled values from weighted food diaries, by food types, Table S3: Cohen Kappa, uniform Kappa, sensitivity and specificity of MFR classification compared to controlled values from weighted food diaries, by food categories, Table S4: MFR portion size estimation performance: mean estimated weights versus true weight , mean error, and mean absolute error of all exactly classified segments, displayed by food types, Table S5: MFR portion size estimation performance: mean estimated weights versus true weights, mean error, and mean absolute error of all exactly classified segments, displayed by food categories, Figure S2: Linear regression of MFR versus controlled values for all found food records, for content of (a) energy; (b) fat; (c) carbohydrates; (d) protein; (e) fiber and (f) alcohol. (Black line: linear regression line; dotted line: 95% limits of agreement; grey line: y=x)., Figure S3: Linear regression of MFR versus controlled values for all found beverage records, for content of (a) energy; (b) fat; (c) carbohydrates; (d) protein; (e) fiber; (f) alcohol. (Black line: linear regression line; dotted line: 95% limits of agreement; grey line: y=x).

Author Contributions: Conceptualization: S.G.N.; methodology: S.G.N., A.C, F.B., C.Z., P.T., G.B.; formal analysis: C.Z., G.B., P.T., T.N.; investigation: C.Z., C.C., W.B.A.; resources: M.B., S.G.N., M.P.L.; data curation: C.Z., G.B.; writing – original draft: C.Z., writing – review and editing: all authors; supervision: S.G.N., M.B., P.T., T.N.; project administration: S.G.N., C.Z.; funding acquisition: S.G.N.

Funding: This research was funded by the CPSLA - Committee for health promotion and fight against addictions of the Canton of Vaud (*Commission de promotion de la santé et de lutte contre les addictions*), via a research grant attributed to S. G. N. The funder had no role in the study design, data collection and analysis, nor in the article writing. F.B. was supported by Swiss Cancer Research (KFS-4722-02-2019).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Thompson, F.E.S., A.F., *Dietary assessment methodology*, in *Nutrition in the Prevention and Treatment of Disease*, A.M. Coulston, Boushey, C.J.; Ferruzzi, M., Editor. 2013, Elsevier: San Diego. p. 5-46.
2. Blanchard, C.M., et al., *Evaluation of PIQNIQ, a Novel Mobile Application for Capturing Dietary Intake.* J Nutr, 2021. **151**(5): p. 1347-1356.
3. Evenepoel, C., et al., *Accuracy of Nutrient Calculations Using the Consumer-Focused Online App MyFitnessPal: Validation Study.* J Med Internet Res, 2020. **22**(10): p. e18237.
4. Illner, A.K., et al., *Review and evaluation of innovative technologies for measuring diet in nutritional epidemiology.* Int J Epidemiol, 2012. **41**(4): p. 1187-203.
5. Sharp, D.B. and M. Allman-Farinelli, *Feasibility and validity of mobile phones to assess dietary intake.* Nutrition, 2014. **30**(11-12): p. 1257-66.
6. Research2Guidance, *mHealth Developer Economics, connectivity in digital health*. 2018, Research2Guidance.
7. Chen, J., J.E. Cade, and M. Allman-Farinelli, *The Most Popular Smartphone Apps for Weight Loss: A Quality Assessment.* JMIR Mhealth Uhealth, 2015. **3**(4): p. e104.
8. Griffiths, C., L. Harnack, and M.A. Pereira, *Assessment of the accuracy of nutrient calculations of five popular nutrition tracking applications.* Public Health Nutr, 2018. **21**(8): p. 1495-1502.
9. Ji, Y., et al., *Validity and Usability of a Smartphone Image-Based Dietary Assessment App Compared to 3-Day Food Diaries in Assessing Dietary Intake Among Canadian Adults: Randomized Controlled Trial.* JMIR Mhealth Uhealth, 2020. **8**(9): p. e16953.
10. Shinozaki, N. and K. Murakami, *Evaluation of the Ability of Diet-Tracking Mobile Applications to Estimate Energy and Nutrient Intake in Japan.* Nutrients, 2020. **12**(11).
11. Marcano-Olivier, M., et al., *Measuring lunchtime consumption in school cafeterias: a validation study of the use of digital photography.* Public Health Nutr, 2019. **22**(10): p. 1745-1754.

12. Gemming, L., et al., *Wearable cameras can reduce dietary under-reporting: doubly labelled water validation of a camera-assisted 24 h recall.* Br J Nutr, 2015. **113**(2): p. 284-91.
13. Jia, W., et al., *Accuracy of food portion size estimation from digital pictures acquired by a chest-worn camera.* Public Health Nutr, 2014. **17**(8): p. 1671-81.
14. Pettitt, C., et al., *A pilot study to determine whether using a lightweight, wearable micro-camera improves dietary assessment accuracy and offers information on macronutrients and eating rate.* Br J Nutr, 2016. **115**(1): p. 160-7.
15. Ahmad, Z., et al., *A Mobile Food Record For Integrated Dietary Assessment.* MADiMa16 (2016), 2016. **2016**: p. 53-62.
16. Casperson, S.L., et al., *A mobile phone food record app to digitally capture dietary intake for adolescents in a free-living environment: usability study.* JMIR Mhealth Uhealth, 2015. **3**(1): p. e30.
17. Rhyner, D., et al., *Carbohydrate Estimation by a Mobile Phone-Based System Versus Self-Estimations of Individuals With Type 1 Diabetes Mellitus: A Comparative Study.* J Med Internet Res, 2016. **18**(5): p. e101.
18. He, H., F. Kong, and J. Tan, *DietCam: Multiview Food Recognition Using a Multikernel SVM.* IEEE J Biomed Health Inform, 2016. **20**(3): p. 848-855.
19. Bucher Della Torre, S., et al., *Development and Evaluation of e-CA, an Electronic Mobile-Based Food Record.* Nutrients, 2017. **9**(1).
20. Lee, C.D., et al., *Comparison of known food weights with image-based portion-size automated estimation and adolescents' self-reported portion size.* J Diabetes Sci Technol, 2012. **6**(2): p. 428-34.
21. Lemacks, J.L., K. Adams, and A. Lovetere, *Dietary Intake Reporting Accuracy of the Bridge2U Mobile Application Food Log Compared to Control Meal and Dietary Recall Methods.* Nutrients, 2019. **11**(1).
22. Wellard-Cole, L., et al., *Relative Validity of the Eat and Track (EaT) Smartphone App for Collection of Dietary Intake Data in 18-to-30-Year Olds.* Nutrients, 2019. **11**(3).
23. Zhang, W., et al., *"Snap-n-Eat": Food Recognition and Nutrition Estimation on a Smartphone.* J Diabetes Sci Technol, 2015. **9**(3): p. 525-33.
24. Höchsmann, C. and C.K. Martin, *Review of the validity and feasibility of image-assisted methods for dietary assessment.* Int J Obes (Lond), 2020. **44**(12): p. 2358-2371.
25. Digital Epidemiology Lab. *The app MyFoodRepo*. Available from: https://www.foodandyou.ch/en/my-food-repo. Accessed May 5, 2021.
26. Office fédéral de la securité alimentaire et des affaires vétérinaires. *Base de données suisse des valeurs nutritives*. Available from: https://valeursnutritives.ch/fr/. Accessed August 17, 2020.
27. Agence nationale de sécurité sanitaire de l'alimentation de l'environnement et du travail (Anses). *Table de composition nutritionnelle des aliments CIQUAL*. Available from: https://ciqual.anses.fr/. Accessed August 17, 2020.
28. Max Rubner-Institut, *German Nutrient Database (Bundeslebensmittelschlüssel, BLS)*. 2020.
29. Société Suisse de Nutrition SSN, *La pyramide alimentaire suisse*, Office fédéral de la securité alimentaire et des affaires vétérinaires, Editor. 2011.
30. Laboratoire d'epidémiologie numérique EPFL. *Open Food Repo*. 2016-2021; Available from: www.foodrepo.org.
31. Brennan, R.L. and D.J. Prediger, *Coefficient Kappa: Some Uses, Misuses, and Alternatives.* Educational and Psychological Measurement, 1981. **41**(3): p. 687-699.
32. Taffe, P., et al., *The Bland-Altman method should not be used when one of the two measurement methods has negligible measurement errors.* Submitted, 2021.

33. Taffe, P., *Effective plots to assess bias and precision in method comparison studies.* Stat Methods Med Res, 2018. **27**(6): p. 1650-1660.
34. Martin, C.K., S. Kaya, and B.K. Gunturk, *Quantification of food intake using food image analysis.* Annu Int Conf IEEE Eng Med Biol Soc, 2009. **2009**: p. 6869-72.
35. Martin, C.K., et al., *Validity of the Remote Food Photography Method (RFPM) for estimating energy and nutrient intake in near real-time.* Obesity (Silver Spring), 2012. **20**(4): p. 891-9.
36. Daugherty, B.L., et al., *Novel technologies for assessing dietary intake: evaluating the usability of a mobile telephone food record among adults and adolescents.* J Med Internet Res, 2012. **14**(2): p. e58.
37. Harnack, L., et al., *Accuracy of estimation of large food portions.* J Am Diet Assoc, 2004. **104**(5): p. 804-6.
38. Flax, V.L., et al., *Validation of a digitally displayed photographic food portion-size estimation aid among women in urban and rural Malawi.* Public Health Nutr, 2019. **22**(17): p. 3140-3150.
39. Norman, Å., et al., *"Everyone can take photos." Feasibility and relative validity of phone photography-based assessment of children's diets - a mixed methods study.* Nutr J, 2020. **19**(1): p. 50.
40. Subar, A.F., et al., *Addressing Current Criticism Regarding the Value of Self-Report Dietary Data.* J Nutr, 2015. **145**(12): p. 2639-45.
41. Khanna, N., et al., *An Overview of The Technology Assisted Dietary Assessment Project at Purdue University.* ISM, 2010: p. 290-295.
42. Woo, I., et al., *Automatic portion estimation and visual refinement in mobile dietary assessment.* Proc SPIE Int Soc Opt Eng, 2010. **7533**.