Public Health Nutrition: 21(15), 2744–2752 doi:10.1017/S1368980018001611

Assessing the relative validity of a new, web-based, self-administered 24 h dietary recall in a French-Canadian population

Jacynthe Lafrenière, Catherine Laramée, Julie Robitaille, Benoît Lamarche and Simone Lemieux*
School of Nutrition, Institute of Nutrition and Functional Foods, Laval University, 2440 Hochelaga Blvd, Québec, QC G1V 0A6, Canada

Submitted 19 September 2017: Final revision received 8 May 2018: Accepted 23 May 2018: First published online 6 July 2018

Abstract

Objective: To assess the relative validity of a new, web-based, self-administered 24 h dietary recall, the R24W, for assessment of energy and nutrient intakes among French Canadians.

Design: Each participant completed a 3 d food record (FR) and the R24W on three occasions over a 4-week period. Intakes of energy and of twenty-four selected nutrients assessed by both methods were compared.

Setting: Québec City metropolitan area.

Subjects: Fifty-seven women and fifty men (mean (SD) age: 47·2 (13·3) years).

Results: Equivalent proportions of under-reporters were found with the R24W (15·0 %) and the FR (23·4 %). Mean (SD) energy intake from the R24W was 7·2 % higher than that from the FR (10 857 (3184) kJ/d (2595 (761) kcal/d) *v.* 10 075 (2971) kJ/d (2408 (710) kcal/d); $P < 0·01$). Significant differences in mean nutrient intakes between the R24W and the FR ranged from –54·8 % (i.e. lower value with R24W) for niacin to +40·0 % (i.e. higher value with R24W) for alcohol. Sex- and energy-adjusted deattenuated correlations between the two methods were significant for all nutrients except Zn (range: 0·35–0·72; $P < 0·01$). Cross-classification demonstrated that 40·0 % of participants were classified in the same quartile with both methods, while 40·0 % were classified in the adjacent quartile and only 3·6 % were grossly misclassified (1st *v.* 4th quartile). Analysis of Bland–Altman plots revealed proportional bias between the two assessment methods for 8/24 nutrients.

Conclusions: These data suggest that the R24W presents an acceptable relative validity as compared with the FR for estimating usual dietary intakes in a cohort of French Canadians.

Keywords
Relative validity
Energy intake
Nutrient intakes
24 h dietary recall
Under-reporting

In the past few decades, researchers have started to rethink how food intake should be assessed and interpreted. Historically, FFQ have been predominantly used in large cohort studies[1–3]. According to Kirkpatrick *et al.*[3], up to 64 % of the previous and ongoing Canadian studies rely on FFQ or dietary screeners while only 14 % are using 24 h recalls. However, studies with recovery biomarkers have consistently reported that multiple 24 h recalls describe energy and protein intakes with higher precision than FFQ[4,5]. While multiple 24 h recalls are considered expensive and time-consuming, web technology now opens the way to a new wave of self-administered automatic tools for use in large cohorts[6,7]. However, the validity of these new automated 24 h recalls has to be demonstrated. To be considered valid and reliable, they have to measure what they are meant to measure consistently over time[8]. On the one hand, they should provide an adequate estimation of nutrient intakes and identify deficiencies[9]. On the other hand, they are also supposed to capture usual intakes. Therefore, reported energy intake (rEI) should be consistent with energy needs to sustain normal activities.

Under-reporting is usually described as implausibly low rEI. To be categorized as such, energy intake has to be significantly lower than estimated or measured daily energy expenditure[10]. The use of doubly labelled water is an unbiased way of assessing daily energy expenditure in real-life settings[11]. In a review published in 2001, Hill

*Corresponding author: Email Simone.lemieux@fsaa.ulaval.ca

and Davies[12] revealed that compared with doubly labelled water, usual rEI from any food assessment tool was associated with a certain degree of under-reporting. Nevertheless, it has been suggested that repeated 24 h recalls would be one of the food assessment methods with the lowest rate of under-reporting, ranging from 10 to 20 %[4,5,13–15]. In the absence of nutritional biomarkers, under-reporting is usually assessed as the ratio between rEI and BMR below the lower limit of physical activity level considered plausible. Goldberg *et al.*[10] suggested that when rEI:BMR is below 1·35, this would be indicative of under-reporting while over-reporting would correspond to rEI:BMR above 2·5. As described by Willett *et al.*[16], nutrient intakes can be further adjusted for energy intake to improve diet description and to strengthen the associations with health outcomes.

An ideal gold standard for dietary intake assessment is difficult to find. Some studies use direct observation, but this is possible only in a clinical setting and not representative of usual intakes. Recovery biomarkers such as doubly labelled water for energy intake are also interesting options. However, such biomarkers mirror specific aspects of the diet, but they cannot reflect global dietary patterns[8]. Newly developed techniques are therefore usually compared with an established one to determine if they can produce equivalent results within predetermined limits[17]. This approach refers to relative validity[9]. In studies evaluating relative validity, authors often use similar established statistical approaches[18,19]. Most often, reported macro- and micronutrient intakes from the new tool are compared with reported intakes obtained from a reference tool. It is expected that this reference method demonstrates a good level of validity, although not necessarily providing a perfect assessment of dietary intakes[9,20]. The food record (FR) has been shown to perform reasonably well compared with biological markers, especially when subjects are asked to weigh their foods and report specific recipes corresponding to what they actually ate[21–24]. This method has been favoured in two recent web-based 24 h recall validation studies because of its independence with the new tools in terms of assessment bias[25,26].

The R24W is a new, automated, self-administered, web-based 24 h recall designed to assess nutritional intakes in the French-Canadian population. This tool uses a data collection approach inspired by the automated multiple-pass method from the US Department of Agriculture[27]. A total of 2568 different food items and 687 recipes are available in the R24W[28]. Respondents are guided to recall their previous day's intake, meal by meal. Pictures of up to eight portion sizes are proposed for each food item described by unit and/or volume. Its development has been discussed in detail elsewhere[28]. A first validation study was conducted in a context of fully controlled feeding studies, in which we showed that there was no systematic bias in portion size estimation with the R24W[29]. The aim of the present study was to assess the relative validity of the R24W, for assessment of energy and nutrient intakes among French Canadians, using established statistical validation approaches and intakes from a 3 d FR as reference. We hypothesized that the R24W accurately estimates participants' usual energy and nutrient intakes with fewer than 20 % of under-reporters.

Methods

Population

Seventy-five women and seventy-five men between 18 and 65 years of age from the Québec City metropolitan area were recruited through electronic messages sent to the Laval University community as well as via the electronic newsletter of the research institute that reaches individuals outside the university. Exclusion criteria were pregnancy, lactation and digestive problems causing malabsorption, to avoid any interaction in the analysis of blood biomarkers taken for an upcoming analysis. All women and seventy-two men completed all the study requirements. The protocol was in accordance with the Declaration of Helsinki and was certified by the Laval University Ethics Committee. All subjects signed a consent form prior taking part in the project.

Study protocol and measurements

Participants were invited to an initial visit at the research institute where their body weight, height and body composition were assessed (TANITA body composition analyser BC-418; Tanita Corporation, IL, USA). Then, they received verbal and written instructions by a dietitian on how to fill out the 3 d FR, with the intent to reduce social desirability biases[30]. They had to complete the record on a weekend day and on two weekdays and they were asked to weigh and measure what they ate as well as to attach recipes or food labels of items consumed to improve accuracy of food assessment. Every FR was revised by a trained dietitian upon return to ensure that the information provided was complete and clear. This was done in order to minimize estimation and reporting errors, since the FR was being used as the reference method in this validation study. Coding was also conducted by trained staff with Nutrific software (Laval University, QC, Canada), which was linked to the Canadian Nutrient File database (Health Canada, 2010).

Afterwards, participants received emails on unannounced days inviting them to complete the R24W four times during a 20 d period. If participants did not complete the 24 h recall on the day they received the email, the access was cancelled and another email was sent on another unannounced day. Briefly, R24W is inspired by the automated multiple-pass method of the US Department of Agriculture[27], but opposed to the automated multiple-pass method, the R24W is using a meal-based

https://doi.org/10.1017/S1368980018001611 Published online by Cambridge University Press

approach in the first step. When completing the R24W, the respondent can add an unlimited number of meals or snacks per 24 h period. In terms of data management, R24W allows automatic calculation of different diet quality scores in addition to energy and nutrient intakes. A detailed description of the R24W has been published elsewhere[28]. As there was no schedule imposed for the completion of the R24W, for the purposes of the current analysis, data of subjects who completed two weekdays and one weekend day were gathered for the comparison with the FR (107 participants; fifty-seven women and fifty men). In cases where all four recalls were eligible, we chose the first two weekdays and the first weekend day completed. Mean intakes from the 3 d FR and from the three R24W days were used in the analyses. During the testing period, subjects were asked not to make any noticeable changes in their usual diet. Use of diet supplements was not taken into account for this validation analysis. Each participant also had to complete questionnaires to gather information about medical history (including questions about weight stability) and sociodemographic variables.

Statistical approaches

Mean daily intakes and standard deviations for energy and twenty-four nutrients were assessed with the R24W and the FR. More precisely, carbohydrates, proteins, fat, percentage of energy from carbohydrates, percentage of energy from proteins and percentage of energy from fat, fibre, vitamin A, thiamin, riboflavin, niacin, vitamin B_6, folic acid, vitamin B_{12}, vitamin C, vitamin D, Mg, Zn, Fe, Ca and K were selected because they are recognized as key nutrients in Canada's Food Guide[31]. SFA, Na and alcohol were also assessed because of their importance in the aetiology of metabolic diseases[32,33]. Student's paired t test was used to determine whether there was a significant difference between the two methods in the assessment of each selected nutrient. Then, the strength of the association between reported intakes using the R24W and reported intakes with the FR was assessed for each nutrient with the Pearson correlation coefficient. Analyses were conducted on raw and on deattenuated sex- and energy-adjusted data. The adjustment for energy was calculated using the residual method[16]. The deattenuation was computed using the ratio of within- to between-person variability of each tool and the number of days of data collection, to adjust for day-to-day variation in intakes[19]. Cross-classification (percentage of agreement) and weighted kappa (κ_w) were assessed to determine if both methods tended to classify respondents in the same quartile. Then, Bland–Altman plots were used to assess agreement at an individual level across the range of intakes. Bland–Altman plots show the relationship between the difference and the average of two measures. A significant association demonstrates a proportional bias between these two measures[34]. Lastly, the relative validity outcome of each test was compared with criteria proposed by Lombard *et al.*[20], based on the work of other authors[35–38] and categorized as good, acceptable or poor to provide an overview of the relative validity of all nutrients tested. Relative validity was considered good in each of these situations: deattenuated sex- and energy-adjusted correlation coefficient ≥0·50; classification of ≥50 % of respondents in the same quartile; classification of <10 % of respondents in opposite quartiles; $\kappa_w \geq 0{\cdot}61$; difference between measures from both methods ≤ 10·9 %; non-significant Student's t test ($P \geq 0{\cdot}05$); and non-significant slope in the Bland–Altman plot ($P \geq 0{\cdot}05$). Relative validity could be judged as being acceptable when the deattenuated sex- and energy-adjusted correlation coefficient was between 0·20 and 0·49; when κ_w was between 0·20 and 0·60; and when the difference between measures from both methods was between 11 and 20 %. Finally, relative validity was considered poor when the deattenuated sex- and energy-adjusted correlation coefficient was < 0·20; when <50 % of respondents were classified in the same quartile; when ≥10 % of respondents were classified in opposite quartiles; when κ_w was <0·20; when the difference between measures from both methods was ≥20 %; and when results from Student's t test and the slope from the Bland–Altman plot were significant ($P \leq 0{\cdot}05$). Agreement between tests and overall relative validity were then evaluated by the total of good, acceptable and poor validity scores obtained for each nutrient.

To determine the relative validity of energy intake assessed by the R24W, a comparison between reported intakes and estimated energy needs was conducted to identify under-reporters, adequate reporters and over-reporters. BMR was estimated with the Mifflin–St Jeor equation[39] and under-reporters were classified as individuals with rEI:BMR <1·35 while over-reporters were classified as those with rEI:BMR >2·5[10]. Lastly, to determine if a similar number of under-reporters was identified with the new tool and the FR, the McNemar χ^2 test for paired data was used to compare under-reporters (rEI:BMR <1·35) and non-under-reporters (rEI:BMR ≥1·35) between the two dietary assessment methods.

Log-transformed data were used to improve normality for all variables. Statistical analyses were conducted with the statistical software package SAS version 9.4.

Results

The main characteristics of the participants are presented in Table 1. Mean age of the participants was 47·4 (SD 13·3) years and they had a mean BMI of 25·5 (SD 4·4) kg/m^2. Fifty-seven per cent of them reported being weight stable for the last 3 months. Ninety-six per cent of participants were Caucasian and 63·6 % had a university degree.

Table 2 presents percentage differences as well as correlations between R24W and FR for energy and nutrient

Table 1 Participants' characteristics: men and women (n 107) aged 18–65 years from the Québec City metropolitan area, Canada. Data collected between March and July 2015

	Mean	SD	%
Women	–	–	53·3
Age (years)	47·4	13·3	–
BMI (kg/m^2)	25·5	4·4	–
Waist circumference (cm)	89·1	13·4	–
Weight stable over the last 3 months	–	–	57·0
Weight gain over the last 3 months	–	–	22·4
Weight loss over the last 3 months	–	–	20·6
Estimated BMR, Mifflin (kJ/d)	6208·2	1228·0	–
Estimated BMR, Mifflin (kcal/d)	1483·8	293·5	–
Ethnicity, Caucasian	–	–	96·3
Education, high school	–	–	5·6
Education, college	–	–	30·8
Education, university	–	–	63·6

intakes. Mean values of eighteen out of twenty-five variables assessed with R24W (72 %) were within 10 % of the mean values obtained with FR. The largest differences were observed for niacin (−54·8 %) and alcohol (+40·0 %). rEI, fat intake, alcohol intake, percentage of energy from carbohydrates and proteins, SFA intake as well as intakes of eight micronutrients were significantly different between the R24W and the FR ($P<0·05$). However, all raw correlations ($r=0·28$–$0·61$) and all but one (Zn at $r=0·02$) sex- and energy-adjusted deattenuated correlations ($r=0·35$–$0·72$) were significant ($P<0·01$).

The cross-classification analysis indicated that, on average, the participants were classified in the same quartile in 40·0 % of the cases (range: 29·9–50·5 %) and in the adjacent quartile in 40·0 % of the cases (range: 27·1–46·7 %), while they were grossly misclassified (e.g. classified in quartile 1 with one method and quartile 4 with the other method) in 3·6 % of the cases (range: 0·9–6·5 %; Table 3). The κ_w values ranged from 0·16 to 0·47 with a mean of 0·33 (Table 3). The Bland–Altman analysis showed a proportional bias for some of the nutrients, but with different patterns. For fat, alcohol, vitamin D and Zn, intakes assessed with the R24W were on average higher than intakes from the FR and the degree of overestimation was proportional to levels of intake. For vitamin A and Mg, there was a noticeable difference in intakes assessed by both tools only in those who reported consuming the largest amounts of these nutrients. Intakes of niacin were underestimated by the R24W compared with the FR and this underestimation became more important in those who consumed a larger amount of niacin. Finally, the intake of vitamin C seemed to be overestimated by the R24W in those who consumed a smaller amount and underestimated in those who consumed a larger amount relatively to the FR (see plots in the online supplementary material). Next, Table 4 combines the relative validation assessment of the six tests performed. Protein was the nutrient for which assessment with both tools demonstrated the highest agreement, while all tests resulted in good or acceptable relative validity outcomes. Carbohydrate, percentage of energy from fat, folic acid, vitamin C, Fe, K and fibre also received mostly results of good or acceptable relative validity outcomes and had only one poor outcome which was related to the proportion of classification in the same quartile (below 50 %). However, for niacin, vitamin C and Zn, results for the majority of the tests (4/7) corresponded to poor outcomes.

Lastly, based on data from the R24W, 15·0 % of participants were characterized as under-reporters, compared with 23·4 % with the FR (Table 5). When we classified the participants as under-reporters or non-under-reporters, we observed that the difference in the proportion of under-reporters between methods did not reach statistical significance ($P=0·07$). Almost three out of every four participants (72·9 %) were classified within the same category by both tools, 26·2 % were one category apart (e.g. under-reporter with one method and adequate reporter with the other one) while only one participant (0·9 %) was grossly misclassified (identified as an under-reporter with the R24W and as an over-reporter with the FR). Lastly, the proportion of participants who reported a recent weight loss was not higher in the under-reporter group (18·8 % in under-reporters *v.* 22·0 % in adequate reporters and 0 % in over-reporters, $P=0·79$, as assessed by the R24W; 28·0 % in under-reporters *v.* 18·8 % in adequate reporters and 0 % in over-reporters, $P=0·48$, as assessed by the FR).

Discussion

It is of first importance to test the validity of newly developed food assessment tools. The present study showed an acceptable level of agreement for energy and nutrient intakes between data generated by a newly developed 24 h recall, the R24W, and data from the 3 d FR, used as the reference method.

In terms of nutrient intakes, our results are comparable to those of the first Belgian food consumption survey using the EPIC-SOFT program, in which intakes from computer-assisted 24 h recalls were compared with intakes assessed by an FR. In that study, raw correlation coefficients between the two methods for energy and nutrient intakes ranged from 0·16 to 0·62[25]. Many nutrients for which significant differences were observed in the first Belgian food consumption survey are the same as the ones for which we observed differences in our study. Indeed, in both studies, there was a difference in reported intakes of energy, fat, SFA, vitamin C, thiamin and riboflavin. Furthermore, results from both studies revealed that energy intake was higher when assessed with 24 h recalls than with the FR, and it was associated with a higher reported intake of fat. As stated by the authors[25], the higher value of reported fat intake with 24 h recalls than with the FR could be related to the numerous questions included about frequently forgotten food items like added fat, spreads or sauces. This higher value of reported fat intake could indeed reflect a more

https://doi.org/10.1017/S1368980018001611 Published online by Cambridge University Press

Table 2 Mean daily intakes of energy and nutrients and correlation coefficients between values derived from the new, web-based, self-administered, 24 h dietary recall (R24W) and the 3 d food record (FR) among men and women (n 107) aged 18–65 years from the Québec City metropolitan area, Canada. Data collected between March and July 2015

	R24W		FR				
	Mean	SD	Mean	SD	% difference	Raw correlation	Sex- and energy-adjusted deattenuated correlation
Energy (kJ)	10 857	3184	10 075	2971	7·2*	0·57*	0·64*
Energy (kcal)	2595	761	2408	710	7·2*	0·57*	0·64*
Carbohydrates (g)	290·6	76·1	277·7	82·7	4·4	0·53*	0·61*
Fat (g)	105·5	44·2	95·8	36·7	9·2*	0·54*	0·54*
Proteins (g)	104·3	32·8	99·7	29·1	4·4	0·61*	0·54*
%E from carbohydrates	43·8	6·9	46·7	7·4	− 6·6*	0·52*	0·62*
%E from fat	35·8	6·0	35·3	6·4	1·4	0·48*	0·63*
%E from proteins	16·2	2·6	16·8	2·8	− 3·7*	0·45*	0·64*
Fibre (g)	25·3	8·7	26·9	8·7	− 6·3	0·47*	0·64*
Vitamin A (μg)	1019·4	726·7	1053·5	982·1	− 3·3	0·35*	0·41*
Thiamin (mg)	2·0	0·6	1·9	0·7	5·0*	0·45*	0·48*
Riboflavin (mg)	2·7	0·9	2·4	0·8	11·1*	0·55*	0·55*
Niacin (mg)	30·5	10·9	47·2	15·5	− 54·8*	0·53*	0·51*
Vitamin B_6 (mg)	2·2	0·6	2·1	0·7	4·5*	0·46*	0·44*
Folic acid (μg)	440·9	130·6	456·8	136·2	− 3·6	0·33*	0·35*
Vitamin B_{12} (μg)	6·1	4·5	5·6	4·4	8·2	0·28*	0·38*
Vitamin C (mg)	131·8	65·6	174·9	96·5	− 32·7*	0·61*	0·72*
Vitamin D (μg)	6·3	4·3	5·5	3·1	12·7	0·35*	0·46*
Mg (mg)	463·7	149·5	461·1	177·8	0·6	0·52*	0·65*
Zn (mg)	14·1	4·8	12·8	3·9	9·2*	0·38*	0·02
Fe (mg)	17·2	4·9	16·5	4·9	4·1	0·34*	0·46*
Ca (mg)	1281·4	450·3	1117·3	396·7	12·8*	0·53*	0·50*
K (mg)	3676·4	954·7	3776·9	990·2	− 2·7	0·53*	0·66*
Alcohol (g)	16·1	15·6	11·5	13·0	40·0*	0·53*	0·69*
SFA (g)	35·4	17·8	30·7	15·1	13·3*	0·58*	0·47*
Na (mg)	3455·4	1127·0	3154·9	1110·0	8·7*	0·55*	0·36*
Mean					1·7	0·48	0·52
SD					16·9	0·09	0·15

%E, percentage of energy.
*Student's t test and Pearson correlation with a P value of <0·05.

reliable assessment of fat, a nutrient known to be often underestimated from biomarker studies[40].

Our analysis showed that the mean sex- and energy-adjusted deattenuated correlation coefficient was 0·52, which respects the criterion for a good relative validity outcome[35]. Regarding cross-classification, although all nutrients except one (protein) did not reach the criterion for good relative validity, our results are comparable to those of others. Indeed, for all nutrients, an average of 80 % of participants were classified in the same or the adjacent quartile. Moreover, fewer than 10 % (range: 0·9–6·5 %) of the participants were classified in the opposite quartile, showing a very low proportion of extreme misclassification. These results are similar to those of a study in which an FFQ was validated with an FR in a similar population where, on average, 77·0 % of participants were classified in the same or the adjacent quartile and 5·0 % were grossly misclassified (opposite quartiles)[41]. The Bland–Altman analysis revealed that the magnitude of the difference between both tools was not equal through the range of mean intakes for eight nutrients. This means that the mean difference between the two tools increases in the larger or the smaller values. This is not an unusual observation. In a study aiming to evaluate the validity of a new FFQ designed for assessing adolescents' intakes, Ambrosini *et al.*[42] observed that 19/22 nutrients tested showed a significant proportional bias in either boys or girls as illustrated by the regression line of the Bland–Altman plot.

The relative validity was not the same for all nutrients studied. However, for fibre, SFA and Na, which are nutrients frequently associated with metabolic health[32,43], we mostly obtained results associated with good or acceptable relative validity. This suggests that the R24W would be an adequate tool to assess dietary intakes in nutritional epidemiological studies addressing issues related to metabolic health. It is worth mentioning that reported intakes for SFA, Na and alcohol are higher with the R24W than with the FR. This supports the idea that social desirability bias is reduced with the web-based dietary assessment tool.

Overall, there are three nutrients in our study for which the relative validity is questionable. For niacin and vitamin C, poor validity outcomes are mainly related to criteria of agreement at a group level; while for Zn, associations and agreement at the individual level as well as agreement at a group level seem to be poor. Since each self-reported dietary assessment tool has some limitations, it is not

Table 3 Cross-classification of daily energy and nutrient intakes into quartiles of the distribution using either the new, web-based, self-administered 24 h dietary recall or the 3 d food record among men and women (*n* 107) aged 18–65 years from the Québec City metropolitan area, Canada. Data collected between March and July 2015

	Same quartile (%)	Adjacent quartiles (%)	±1 quartile apart (%)	Misclassified (1st *v.* 4th quartile) (%)	κ_w
Energy (kcal)	40·2	43·0	83·2	3·7	0·35
Carbohydrates (g)	42·1	42·1	84·1	6·5	0·35
Fat (g)	45·8	36·4	82·2	2·8	0·40
Proteins (g)	50·5	38·3	88·8	4·7	0·47
%E from carbohydrates	34·6	42·1	76·6	4·7	0·25
%E from fat	35·5	42·1	77·6	2·8	0·28
%E from proteins	45·8	31·8	77·6	3·7	0·35
Fibre (g)	43·0	42·1	85·0	2·8	0·40
Vitamin A (μg)	47·7	27·1	74·8	4·7	0·34
Thiamin (mg)	31·8	45·8	77·6	2·8	0·25
Riboflavin (mg)	40·2	40·2	80·4	0·9	0·35
Niacin (mg)	41·1	43·0	84·1	1·9	0·38
Vitamin B_6 (mg)	33·6	45·8	79·4	6·5	0·25
Folic acid (μg)	39·3	35·5	74·8	5·6	0·26
Vitamin B_{12} (μg)	38·3	42·1	80·4	6·5	0·29
Vitamin C (mg)	42·1	39·3	81·3	1·9	0·37
Vitamin D (μg)	29·9	41·1	71·0	5·6	0·16
Mg (mg)	42·1	34·6	76·6	4·7	0·31
Zn (mg)	31·8	40·2	72·0	2·8	0·20
Fe (mg)	36·4	42·1	78·5	4·7	0·28
Ca (mg)	39·3	44·9	84·1	1·9	0·37
K (mg)	36·4	46·7	83·2	1·9	0·34
Alcohol (g)	50·5	33·6	84·1	1·9	0·47
SFA (g)	37·4	45·8	83·2	2·8	0·34
Na (mg)	46·7	33·6	80·4	1·9	0·40
Mean	40·0	40·0	80·0	3·6	0·33
SD	5·7	5·0	4·3	1·7	0·10

κ_w, weighted kappa; %E, percentage of energy.

possible to determine based on our results that the R24W would systematically produce erroneous estimation for these specific nutrients. However, it would be wiser to interpret with caution estimation of those nutrients evaluated with the R24W. Our next step will be to identify food items that could explain the large discrepancies between the two methods compared. In a larger perspective, it seems that the tests used to evaluate agreement at a group level (percentage difference, Student's *t* test and Bland–Altman) and those evaluating agreement at an individual level (Pearson correlations, cross-classification and κ_w) were characterized by an equivalent number of good and poor outcomes for the majority of nutrients tested. This, combined with the small proportion of the cohort characterized as under-reporters, suggests that this new, web-based 24 h recall would be suitable to assess dietary intakes in research projects aiming to evaluate intakes at either a group or an individual level.

Under-reporting of dietary intakes has been identified as a major issue for which dietary assessment tools are often criticized. However, the current study demonstrated that the R24W did not produce a higher prevalence of under-reporters compared with the FR. Prentice *et al.* conducted a study where they compared reported energy intake using a 4 d FR and three 24 h recalls with doubly labelled water as a biomarker of energy intake. Similar to what we found, they noticed only a slight difference between the two methods in the proportion of participants identified as under-reporters[5].

It is important to mention that the present study aimed to compare usual intakes as assessed by two different tools and that a perfect agreement was not expected. Indeed, data were collected on different days with two self-reported methods associated with some degree of imprecision. Furthermore, even if the FR is considered a gold standard, it has been widely reported that individuals who fill in FR tend to modify what they eat because they know they are being evaluated. This is called reactivity bias. It could result in an underestimation of some nutrients such as fat and alcohol[44] and in an apparent overestimation generated by other tools in a comparative context[19]. This could explain the discrepancies observed between the FR and the R24W for these two nutrients. The reactivity bias is not a problem with a 24 h recall because participants do not know in advance which days will be assessed. However, if participants experience difficulties with short-term memory, assessment by the 24 h recall would be affected[45]. Moreover, the R24W offers a wide selection of food items and mixed dishes[28] but, contrary to the FR where participants could write virtually any possible item, in the R24W choices are limited, which could force some respondents to use predetermined recipes slightly different from what they actually ate.

We also stress that we conducted the current study with a rather small homogeneous cohort of highly educated adults that is not fully representative of the French-Canadian population. These characteristics of the sample limit the generalizability of the results to different populations.

Table 4 Statistical test outcomes and proportion of poor outcomes for the relative validity of the new, web-based, self-administered 24 h dietary recall compared with the 3 d food record among men and women (*n* 107) aged 18–65 years from the Québec City metropolitan area, Canada. Data collected between March and July 2015

	Validity at the individual level			Validity at a group level			
Characteristics assessed	Association	Agreement	Agreement	Agreement	Agreement	Presence, direction and extent of bias	
Tests	Correlation coefficient	Cross-classification	κ_w	% difference	Student's *t* test	Bland–Altman (slope of the regression)	Proportion of poor outcomes (/7)
Criteria for a good outcome (G)	≥0·50	≥50% in same quartile; <10% in opposite quartile	≥0·61	0–10·9%	$P>0·05$	$P>0·05$	
Criteria for an acceptable outcome (A)	0·20–0·49		0·20–0·60	11·0–20 %			
Criteria for a poor outcome (P)	<0·20	<50 % in same quartile; ≥10 % in opposite quartile	<0·20	>20%	$P\leq 0·05$	$P\leq 0·05$	
Energy (kcal)	G	P–G	A	G	P	G	2
Carbohydrates (g)	G	P–G	A	G	G	G	1
Fat (g)	G	P–G	A	G	P	P	3
Proteins (g)	G	G–G	A	G	G	G	0
%E from carbohydrates	G	P–G	A	G	P	G	2
%E from fat	G	P–G	A	G	G	G	1
%E from proteins	G	P–G	A	A	P	G	2
Fibre (g)	G	P–G	A	G	G	G	1
Vitamin A (μg)	A	P–G	A	G	G	P	2
Thiamin (mg)	A	P–G	A	G	P	G	2
Riboflavin (mg)	G	P–G	A	A	P	G	2
Niacin (mg)	G	P–G	A	P	P	P	4
Vitamin B_6 (mg)	A	P–G	A	G	P	G	2
Folic acid (μg)	A	P–G	A	G	G	G	1
Vitamin B_{12} (μg)	A	P–G	A	G	G	G	1
Vitamin C (mg)	G	P–G	A	P	P	P	4
Vitamin D (μg)	A	P–G	P	A	G	P	3
Mg (mg)	G	P–G	A	G	G	P	2
Zn (mg)	P	P–G	A	G	P	P	4
Fe (mg)	A	P–G	A	G	G	G	1
Ca (mg)	G	P–G	A	A	P	G	2
K (mg)	G	P–G	A	G	G	G	1
Alcohol (g)	G	G–G	A	P	P	P	3
SFA (g)	A	P–G	A	G	P	G	2
Na (mg)	A	P–G	A	G	P	G	2

κ_w, weighted kappa; %E, percentage of energy.

Table 5 Proportion of under-, adequate and over-reporters, assessed by the ratio between reported energy intake and BMR (rEI:BMR), as determined with the new, web-based, self-administered 24 h dietary recall (R24W) and the 3 d food record (FR) among men and women (*n* 107) aged 18–65 years from the Québec City metropolitan area, Canada. Data collected between March and July 2015

	Total (*n* 107)		Men (*n* 50)		Women (*n* 57)	
	R24W (%)	FR (%)	R24W (%)	FR (%)	R24W (%)	FR (%)
Under-reporters (rEI:BMR <1·35)	15·0	23·4	20·0	24·0	10·5	22·8
Adequate reporters (rEI:BMR = 1·35–2·5)	80·4	74·8	72·0	72·0	87·7	77·2
Over-reporters (rEI:BMR >2·5)	4·7	1·9	8·0	4·0	1·7	0·0

Furthermore, we only used three days of FR and of R24W. For the purposes of the present study, we stipulated that this period represented a good estimation of usual intakes. We decided to do so to limit the burden on participants and also because we wanted to validate the tool in a context suitable for larger studies. It is also of importance to mention that we decided not to exclude the under-reporters from the analysis to keep a representative sample.

Compared with most of the validation studies published so far, we improved the analysis by pooling the results of six validation tests to get an overview of the validity for each nutrient. This approach allowed us to identify for which nutrient the tool was more effective, using the FR as a reference method.

Conclusion

The present paper assessed the relative validity of a new, web-based, self-administered 24 h recall, the R24W, for intakes of energy and twenty-four selected nutrients using six different statistical tests in a cohort of French-Canadian adults. This comparative analysis with the FR suggests that the R24W has an acceptable level of relative validity for most nutrients as well as for energy. However, assessment of niacin, vitamin C and Zn with the R24W should be interpreted with caution considering results obtained in the present study.

Acknowledgements

Acknowledgements: The authors would like to acknowledge the work of Pascale Bélanger, Myriam Landry, Amélie Bergeron and Caroline Trahan in the coding of food records. *Financial support:* This work was funded by the Canadian Institutes of Health Research (grant no. FHG 129921). The Canadian Institutes of Health Research had no role in the design, analysis or writing of this article. *Conflicts of interest:* None. *Authorship:* J.L., J.R., B.L. and S.L. were involved in formulating the research question and designing the study. C.L. was involved in carrying out the study J.L. was in charge of analysing the data and writing the article. All authors reviewed and approved the final version. *Ethics of human subject participation:* This study was conducted according to the guidelines laid down in the Declaration of Helsinki and all procedures involving human subjects were approved by the Laval University Ethics Committee. Written informed consent was obtained from all subjects.

Supplementary material

To view supplementary material for this article, please visit https://doi.org/10.1017/S1368980018001611

References

1. Shim J-S, Shin H-R, Oh K *et al.* (2014) Dietary assessment methods in epidemiologic studies. *Epidemiol Health* **36**, e2014009.
2. Winpenny EM, Penney TL, Corder K *et al.* (2017) Change in diet in the period from adolescence to early adulthood: a systematic scoping review of longitudinal studies. *Int J Behav Nutr Phys Act* **14**, 60.
3. Kirkpatrick SI, Vanderlee L, Raffoul A *et al.* (2017) Self-report dietary assessment tools used in Canadian research: a scoping review. *Adv Nutr* **8**, 276–289.
4. Freedman LS, Commins JM, Moler JE *et al.* (2014) Pooled results from 5 validation studies of dietary self-report instruments using recovery biomarkers for energy and protein intake. *Am J Epidemiol* **180**, 172–188.
5. Prentice RL, Mossavar-Rahmani Y, Huang Y *et al.* (2011) Evaluation and comparison of food records, recalls, and frequencies for energy and protein assessment by using recovery biomarkers. *Am J Epidemiol* **174**, 591–603.
6. Kohlmeier L, Mendez M, McDuffie J *et al.* (1997) Computer-assisted self-interviewing: a multimedia approach to dietary assessment. *Am J Clin Nutr* **65**, 4 Suppl., 1275S–1281S.
7. Taren D, Dwyer J, Freedman L *et al.* (2002) Dietary assessment methods: where do we go from here? *Public Health Nutr* **5**, 1001–1003.
8. Margetts BM & Nelson M (1997) *Design Concepts in Nutritional Epidemiology*. Oxford: Oxford University Press.
9. Gleason PM, Harris J, Sheean PM *et al.* (2010) Publishing nutrition research: validity, reliability, and diagnostic test assessment in nutrition-related research. *J Am Diet Assoc* **110**, 409–419.
10. Goldberg GR & Black AE (1998) Assessment of the validity of reported energy intakes – review and recent developments. *Food Nutr Res* **42**, 6–9.
11. Schoeller DA (1995) Limitations in the assessment of dietary energy intake by self-report. *Metabolism* **44**, 18–22.
12. Hill RJ & Davies PSW (2001) The validity of self-reported energy intake as determined using the doubly labelled water technique. *Br J Nutr* **85**, 415–430.
13. Subar AF, Kipnis V, Troiano RP *et al.* (2003) Using intake biomarkers to evaluate the extent of dietary misreporting in a large sample of adults: the OPEN study. *Am J Epidemiol* **158**, 1–13.
14. Schatzkin A, Kipnis V, Carroll RJ *et al.* (2003) A comparison of a food frequency questionnaire with a 24-hour recall for use in an epidemiological cohort study: results from the biomarker-based Observing Protein and Energy Nutrition (OPEN) study. *Int J Epidemiol* **32**, 1054–1062.
15. Tooze JA, Subar AF, Thompson FE *et al.* (2004) Psychosocial predictors of energy underreporting in a large doubly labeled water study. *Am J Clin Nutr* **79**, 795–804.
16. Willett WC, Howe GR & Kushi LH (1997) Adjustment for total energy intake in epidemiologic studies. *Am J Clin Nutr* **65**, 4 Suppl., 1220S–1228S.
17. Hanneman SK (2008) Design, analysis and interpretation of method-comparison studies. *AACN Adv Crit Care* **19**, 223–234.
18. Cade JE, Burley VJ, Warm DL *et al.* (2004) Food-frequency questionnaires: a review of their design, validation and utilisation. *Nutr Res Rev* **17**, 5–22.
19. Willett WC (1998) *Nutritional Epidemiology*. New York: Oxford University Press.
20. Lombard MJ, Steyn NP, Charlton KE *et al.* (2015) Application and interpretation of multiple statistical tests to evaluate validity of dietary intake assessment methods. *Nutr J* **14**, 40.
21. Barrett JS & Gibson PR (2010) Development and validation of a comprehensive semi-quantitative food frequency questionnaire that includes FODMAP intake and glycemic index. *J Am Diet Assoc* **110**, 1469–1476.
22. Bingham SA, Cassidy A, Cole TJ *et al.* (1995) Validation of weighed records and other methods of dietary assessment using the 24 h urine nitrogen technique and other biological markers. *Br J Nutr* **73**, 531–550.
23. Fitt E, Cole D, Ziauddeen N *et al.* (2015) DINO (Diet In Nutrients Out) – an integrated dietary assessment system. *Public Health Nutr* **18**, 234–241.
24. Ortega RM, Pérez-Rodrigo C & López-Sobaler AM (2015) Dietary assessment methods: dietary records. *Nutr Hosp* **31**, Suppl. 3, 38–45.
25. De Keyzer W, Huybrechts I, De Vriendt V *et al.* (2011) Repeated 24-hour recalls versus dietary records for estimating

nutrient intakes in a national food consumption survey. *Food Nutr Res* **2011**, 55.
26. Comrie F, Masson LF & McNeill G (2009) A novel online Food Recall Checklist for use in an undergraduate student population: a comparison with diet diaries. *Nutr J* **8**, 13.
27. Moshfegh AJ, Rhodes DG, Baer DJ *et al.* (2008) The US Department of Agriculture Automated Multiple-Pass Method reduces bias in the collection of energy intakes. *Am J Clin Nutr* **88**, 324–332.
28. Jacques S, Lemieux S, Lamarche B *et al.* (2016) Development of a web-based 24-h dietary recall for a French-Canadian population. *Nutrients* **8**, 724.
29. Lafrenière J, Lamarche B, Laramée C *et al.* (2017) Validation of a newly automated web-based 24-hour dietary recall using fully controlled feeding studies. *BMC Nutr* **3**, 34.
30. Turgeon O'Brien H & Dufour R (1994) Calcium and magnesium intakes during pregnancy: association with blood pressure. *Top Clin Nutr* **9**, 74–85.
31. Health Canada (2011) *Eating Well with Canada's Food Guide – A Resource for Educators and Communicators*. Ottawa, ON: Health Canada; available at http://www.healthcanada.gc.ca/foodguide
32. Eckel RH, Jakicic JM, Ard JD *et al.* (2014) 2013 AHA/ACC guideline on lifestyle management to reduce cardiovascular risk: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *J Am Coll Cardiol* **63**, 2960–2984.
33. Corrao G, Rubbiati L, Bagnardi V *et al.* (2000) Alcohol and coronary heart disease: a meta-analysis. *Addiction* **95**, 1505–1523.
34. Bland JM & Altman DG (1999) Measuring agreement in method comparison studies. *Stat Methods Med Res* **8**, 135–160.
35. Masson LF, McNeill G, Tomany JO *et al.* (2003) Statistical approaches for assessing the relative validity of a food-frequency questionnaire: use of correlation coefficients and the kappa statistic. *Public Health Nutr* **6**, 313–321.
36. Cade J, Thompson R, Burley V *et al.* (2002) Development, validation and utilisation of food-frequency questionnaires – a review. *Public Health Nutr* **5**, 567–587.
37. Luevano-Contreras C, Durkin T, Pauls M *et al.* (2013) Development, relative validity, and reliability of a food frequency questionnaire for a case–control study on dietary advanced glycation end products and diabetes complications. *Int J Food Sci Nutr* **64**, 1030–1035.
38. Willett WC, Sampson L, Stampfer MJ *et al.* (1985) Reproducibility and validity of a semiquantitative food frequency questionnaire. *Am J Epidemiol* **122**, 51–65.
39. Frankenfield D, Roth-Yousey L & Compher C (2005) Comparison of predictive equations for resting metabolic rate in healthy nonobese and obese adults: a systematic review. *J Am Diet Assoc* **105**, 775–789.
40. Heitmann BL, Lissner L & Osler M (2000) Do we eat less fat, or just report so? *Int J Obes Relat Metab Disord* **24**, 435–442.
41. Labonté M-È, Cyr A, Baril-Gravel L *et al.* (2012) Validity and reproducibility of a web-based, self-administered food frequency questionnaire. *Eur J Clin Nutr* **66**, 166–173.
42. Ambrosini GL, de Klerk NH, O'Sullivan TA *et al.* (2009) The reliability of a food frequency questionnaire for use among adolescents. *Eur J Clin Nutr* **63**, 1251–1259.
43. Chiuve SE, Fung TT, Rimm EB *et al.* (2012) Alternative dietary indices both strongly predict risk of chronic disease. *J Nutr* **142**, 1009–1018.
44. Midanik LT (1988) Validity of self-reported alcohol use: a literature review and assessment. *Br J Addict* **83**, 1019–1029.
45. National Institutes of Health National Cancer Institute (n.d.) Dietary Assessment Primer | 24-hour Dietary Recall. http://dietassessmentprimer.cancer.gov/profiles/ (accessed March 2016).