**Title**: Assessing the relative validity of a new web-based self-administered 24-hour dietary recall in a French-Canadian Population.

This paper was published in Public Health Nutrition May 23rd, 2018

**Authors** : Jacynthe Lafrenière[1], Catherine Laramée[1], Julie Robitaille[1], Benoît Lamarche[1], Simone Lemieux[1]

1. Institute of Nutrition and Functional Foods, Laval University
2440 Hochelaga Blvd, Quebec, Qc Canada

**Corresponding Author**: Simone Lemieux, Dt.P., Ph.D.
Simone.lemieux@fsaa.ulaval.ca
School of Nutrition
Laval University – Institute of Nutrition and Functional Foods
2440 Hochelaga Blvd, Quebec, Qc Canada
418-656-2131 ext 3637

**Short Title**: Validity of a new 24-hour dietary recall

<u>Disclosures</u>

**Acknowledgments**: 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 Canadian Institutes of Health Research (grant no. FHG 129921)

**Conflicts of Interest**: None

Abbreviations:

BMR: Basal metabolic rate
DEE: Daily energy expenditure
DLW: Doubly labeled water
FR: Food record
FFQ: Food frequency questionnaire
rEI: Reported energy intake

**Abstract:**

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

Design: Each participant completed a 3-day food record (FR) and the R24W on three occasions over a 4-week period. Intakes of energy and of 24 selected nutrients assessed by both methods were compared.

Setting: Quebec City metropolitan area.

Subjects: Fifty-seven women and 50 men (mean 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 energy intake derived from the R24W was 7.2% higher than the value obtained with the FR (2595±761 vs. 2408±710 kcal/d; $P<0.01$). Significant differences in mean intakes of nutrients 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 de-attenuated correlations between the two methods were significant for all nutrients except zinc and ranged from 0.35 to 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 (1$^{st}$ vs. 4$^{th}$ quartile). Analysis of Bland-Altman plots revealed proportional bias between the two assessment methods for 8/24 nutrients.

Conclusion: These data suggest that this new web-based self-administered 24-hour dietary recall presents an acceptable relative validity as compared with FR for estimating usual dietary intakes in a cohort of French Canadians.

**Keywords:** Relative validity, energy intake, nutrient intakes, 24-hour dietary recall, under-reporting.

## Introduction:

In the past few decades, researchers have started to rethink how food intake should be assessed and interpreted. Historically, food frequency questionnaires (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-hour recalls. However, studies with recovery biomarkers have consistently reported that multiple 24-hour recalls described energy and protein intakes with higher precision than FFQ (4,5). While multiple 24-hour 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-hour 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 (DEE)(10). The use of doubly labeled water (DLW) is an unbiased way of assessing DEE in real life settings (11). In a review published in 2001, Hill et al. (12) revealed that compared to DLW, usual rEI from any food assessment tool was associated with a certain degree of under-reporting. Nevertheless, it has been suggested that repeated 24-hour recalls would be one of the food assessment method 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 basal metabolic rate (BMR) considered as below the lower limit of plausible physical activity level. Goldberg et al. (10) suggested that when a rEI:BMR ratio is below 1.35, this would be indicative of under-reporting while over-reporting would correspond to a rEI:BMR ratio above 2.5. As described by Willett et al. (16), nutrient intakes can be further adjusted for energy intake in order to improve diet description and to strengthen the associations with health outcomes.

Ideal gold standard for dietary intake assessment is difficult to find. Some studies use direct observation, but this is only possible in a clinical setting and not representative of usual intakes. Recovery biomarkers such as DLW 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 used 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 when compared to 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-hour recall validation studies because of its independence with the new tools in term of assessment bias (25,26).

The R24W is a new automated, self-administered web-based 24-hour 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 (AMPM) from the United States Department of Agriculture (USDA) (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 have shown 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 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 75 men between the age of 18 and 65 from the Quebec 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 of the university. Exclusion criteria were pregnancy, lactation and digestive problems causing malabsorption in order to avoid any interaction in the analysis of blood biomarkers taken for an upcoming analysis. All women and 72 men completed all the study requirements. [Ethics]

### 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 analyzer BC-418, Tanita Corporation, Il, USA). Then, they received verbal and written instructions by a dietitian on how to fill out the 3-day 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 information provided was complete and clear. This was done in order to minimize estimation and reporting errors, the FR being used as the reference method in this validation study. Coding was also conducted by a trained staff with Nutrific software (Laval University, Qc, Canada), which was linked to the Canadian Nutrient File database (Health Canada, 2010).

Afterwards, participants received e-mails on unannounced days inviting them to complete the R24W four times during a 20-day period. If participants did not complete the 24-hour recall on the day they received the e-mail, the access was cancelled and another e-mail was sent on another unannounced day. Briefly, R24W is inspired by the AMPM of the USDA (27), but as opposed to the AMPM, R24W is using a meal-based approach in the first step.

When completing the R24W, the respondent can add an unlimited number of meals or snacks per 24-hour 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 purpose of this analysis, data of subjects who completed two weekdays and one weekend day were gathered for the comparison with the FR (107 participants; 57 women and 50 men). In cases where all 4 recalls were eligible, we chose the first two weekdays and the first weekend day completed. Mean intakes from the 3-day FR and from the three days of R24W 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 socio-demographic variables.

## Statistical Approaches

Mean intakes and standard deviations for energy and 24 nutrients were assessed with the R24W and the FR. More precisely, carbohydrates, proteins, fat, % of energy from carbohydrates, % of energy from proteins and % of energy from fat, fibre, vitamin A, thiamin, riboflavin, niacin, vitamin B6, folic acid, vitamin B12, vitamin C, vitamin D, magnesium, zinc, iron, calcium and potassium were selected because they are recognized as key nutrients in the Canada's Food Guide (31). Saturated fatty acids, sodium and alcohol were also assessed because of their importance in the etiology 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 Pearson correlation coefficient. Analyses were conducted on raw and on de-attenuated sex- and energy-adjusted data. The adjustment for energy was calculated using the residual method (16). The de-attenuation was computed using the ratio of within and between person variability of each tool and the number of days of data collection in order to adjust for day-to-day variation in intakes (19). Cross-classification (percentage of agreement) and weighted Kappa were assessed in order 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 relation between the difference and the average of two measures. A significant association demonstrates a proportional bias between these two measures (34). Lastly, 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 (G), acceptable (A) or poor (P) to provide an overview of the relative validity of all nutrients tested. Relative validity was considered good in each of these situations: de-attenuated sex- and energy-adjusted correlation coefficient ≥0.50, classification of ≥50% of respondents in the same quartile, classification of <10% of respondents in the opposite quartiles, weighted Kappa ≥0.61, ≤10.9% of difference between measures from both methods, non-significant (P≥0.05) Student's t-test and non-significant slope in the Bland-Altman plot (P≥0.05). Relative validity could be judged as being acceptable when de-attenuated sex- and energy-adjusted correlation coefficient was between 0.20 and 0.49, when the weighted Kappa 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 de-attenuated 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 the opposite quartile, when weighted Kappa 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 (≤0.05). Agreement between tests and overall relative validity were then evaluated by the total of good, acceptable and poor validity scores obtained for each nutrient.

In order 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 ratio <1.35 while over-reporters were classified as those with rEI:BMR ratio > 2.5 (10). Lastly, to determine if a similar number of under-reporters was identified with the new tool and the FR, the McNemar chi-square 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 software SAS version 9.4 (SAS Institute Inc., NC, USA).

**Results**

The main characteristics of the participants are presented in **Table 1**. Mean age of the participants was 47.4 ±13.3 years and they had a mean body mass index of 25.5 ± 4.4 kg/m$^2$. Fifty-seven percent of them have reported being weight stable for the last three months. Ninety-six percent of participants were Caucasian and 63.6% had a university degree.

**Table 2** presents differences in percentage as well as correlations between R24W and FR for energy and nutrient intakes. Mean values of 18 out of 25 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, % of energy from carbohydrates and proteins, saturated fatty acid 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 (zinc at r=0.02) sex- and energy-adjusted de-attenuated 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**). Kappa statistics ranged from 0.16 to 0.47 with an average 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 zinc, 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 magnesium, there was only a noticeable difference in intakes assessed by both tools in those who reported consuming the largest amounts of these nutrients. Intakes of niacin were underestimated by the R24W compared to 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 (**plots in 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, percent of fat, folic acid, vitamin C, iron, potassium and fibre also received mostly results of good or acceptable relative validity outcomes and only had one poor outcome which was related to the proportion of classification in the same quartile (below 50%). However, for niacin, vitamin C and zinc, 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 (for example, 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 vs. 22.0% in adequate reporters and 0% in over-reporters, P=0.79 as assessed by the R24W; 28.0% in under-reporters vs. 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. This study showed an acceptable level of agreement for energy and nutrient intakes between data generated by a newly developed 24-hour recall, the R24W, and data from the FR, used as the reference method.

In terms of nutrient intakes, our results are comparable to those of the EPIC-soft study, in which intakes from computer-assisted 24-hour recalls were compared to intakes assessed with a 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 EPIC-soft study 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, saturated fatty acids, vitamin C, thiamin and riboflavin. Furthermore, results from both studies revealed that energy intake was higher when assessed with 24-hour recalls than with the FR, and it was associated with a higher reported intake of fat. As stated by De Keyzer et al. (25), the higher value of reported fat intake with 24-hour 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 reliable assessment of fat, a nutrient known to be often underestimated in biomarkers studies (40).

Our analysis showed that the average sex- and energy-adjusted de-attenuated 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 a FFQ was validated with a 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 average 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, saturated fatty acids and sodium, 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 saturated fatty acids, sodium 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 zinc, 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 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 (% difference, Student's t-test and Bland-Altman) and those evaluating agreement at an individual level (Pearson correlations, cross-classification and Kappa) 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-hour recall would be suitable to assess dietary intakes in research projects aiming to evaluate intakes either at a group or at an individual level.

Under-reporting of dietary intakes has been identified as a major issue for which dietary assessment tools are often criticized. However, this study demonstrated that the R24W did not produce a higher prevalence of under-reporters compared to the FR. Prentice et al.

(2011) conducted a study where they compared reported energy intake using a 4-day FR and 3 days of 24-hour recall with DLW 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 this 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 as a gold standard, it has been widely reported that individuals who filled in FR tend to modify what they eat because they know they are being evaluated. This is called a reactivity bias. This 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-hour 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-hour 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 virtually write 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 have to stress that we conducted this study with a rather small homogenous 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. Furthermore, we only used three days of FR and of R24W. For the purpose of this 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 in order to keep a representative sample.

Compared to 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

This paper assessed the relative validity of a new web-based self-administered 24-hour recall, the R24W, for energy intake and 24 selected nutrients using six different statistical tests in a cohort of French-Canadian adults. This comparative analysis with 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 zinc with the R24W should be interpreted with caution considering results obtained in the present study.

## References


1. Shim J-S, Shin H-R, Oh K, Varghese C, Kim HC. Dietary assessment methods in epidemiologic studies. Epidemiol Health. 2014 Jul 22;e2014009.

2. Winpenny EM, Penney TL, Corder K, White M, Sluijs EMF. Change in diet in the period from adolescence to early adulthood: a systematic scoping review of longitudinal studies. Int J Behav Nutr Phys Act. 2017 May 4;14(1):60.

3. Kirkpatrick SI, Vanderlee L, Raffoul A, Stapleton J, Csizmadi I, Boucher BA, et al. Self-Report Dietary Assessment Tools Used in Canadian Research: A Scoping Review. Adv Nutr. 2017 Mar 1;8(2):276–89.

4. Freedman LS, Commins JM, Moler JE, Arab L, Baer DJ, Kipnis V, et al. Pooled Results From 5 Validation Studies of Dietary Self-Report Instruments Using Recovery Biomarkers for Energy and Protein Intake. Am J Epidemiol. 2014 Jul 15;180(2):172–88.

5. Prentice RL, Mossavar-Rahmani Y, Huang Y, Van Horn L, Beresford SAA, Caan B, et al. Evaluation and comparison of food records, recalls, and frequencies for energy and protein assessment by using recovery biomarkers. Am J Epidemiol. 2011 Sep 1;174(5):591–603.

6. Kohlmeier L, Mendez M, McDuffie J, Miller M. Computer-assisted self-interviewing: a multimedia approach to dietary assessment. Am J Clin Nutr. 1997 Jan 4;65(4):1275S-1281S.

7. Taren D, Dwyer J, Freedman L, Solomons NW. Dietary assessment methods: where do we go from here? Public Health Nutr. 2002 Dec;5(6A):1001–3.

8. Margetts BM, Nelson M. Design concepts in nutritional epidemiology [Internet]. Oxford: Oxford University Press; 1997. Available from: Accès via Oxford scholarship online http://dx.doi.org/10.1093/acprof:oso/9780192627391.001.0001

9. Gleason PM, Harris J, Sheean PM, Boushey CJ, Bruemmer B. Publishing Nutrition Research: Validity, Reliability, and Diagnostic Test Assessment in Nutrition-Related Research. J Am Diet Assoc. 2010 Mar 1;110(3):409–19.

10. Goldberg GR, Black AE. Assessment of the validity of reported energy intakes - review and recent developments. Food Nutr Res. 1998 Jun 1;42(0).

11. Schoeller DA. Limitations in the assessment of dietary energy intake by self-report. Metabolism. 1995;44:18–22.

12. Hill RJ, Davies PSW. The validity of self-reported energy intake as determined using the doubly labelled water technique. Br J Nutr. 2001 Apr;85(04):415–430.

13. Subar AF, Kipnis V, Troiano RP, Midthune D, Schoeller DA, Bingham S, et al. Using Intake Biomarkers to Evaluate the Extent of Dietary Misreporting in a Large Sample of Adults: The OPEN Study. Am J Epidemiol. 2003 Jan 7;158(1):1–13.

14. Schatzkin A, Kipnis V, Carroll RJ, Midthune D, Subar AF, Bingham S, et al. 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. 2003 Jan 12;32(6):1054–62.

15. Tooze JA, Subar AF, Thompson FE, Troiano R, Schatzkin A, Kipnis V. Psychosocial predictors of energy underreporting in a large doubly labeled water study. Am J Clin Nutr. 2004 Jan 5;79(5):795–804.

16. Willett WC, Howe GR, Kushi LH. Adjustment for total energy intake in epidemiologic studies. Am J Clin Nutr. 1997 Jan 4;65(4):1220S-1228S.

17. Hanneman SK. Design, Analysis and Interpretation of Method-Comparison Studies. AACN Adv Crit Care. 2008 Jun;19(2):223.

18. Cade JE, Burley VJ, Warm DL, Thompson RL, Margetts BM. Food-frequency questionnaires: a review of their design, validation and utilisation. Nutr Res Rev. 2004 Jun;17(01):5–22.

19. Willett WC. Nutritional Epidemiology [Internet]. Oxford University Press; 1998 [cited 2016 Jan 26]. Available from: http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780195122978.001.0001/acprof-9780195122978

20. Lombard MJ, Steyn NP, Charlton KE, Senekal M. Application and interpretation of multiple statistical tests to evaluate validity of dietary intake assessment methods. Nutr J. 2015 Apr 22;14:40.

21. Barrett JS, Gibson PR. Development and Validation of a Comprehensive Semi-Quantitative Food Frequency Questionnaire that Includes FODMAP Intake and Glycemic Index. J Am Diet Assoc. 2010 Oct;110(10):1469–76.

22. Bingham SA, Cassidy A, Cole TJ, Welch A, Runswick SA, Black AE, et al. Validation of weighed records and other methods of dietary assessment using the 24 h urine nitrogen technique and other biological markers. Br J Nutr. 1995 Apr;73(4):531–50.

23. Fitt E, Cole D, Ziauddeen N, Pell D, Stickley E, Harvey A, et al. DINO (Diet In Nutrients Out) - an integrated dietary assessment system. Public Health Nutr. 2015 Feb;18(2):234–41.

24. Ortega RM, Pérez-Rodrigo C, López-Sobaler AM. Dietary assessment methods: dietary records. Nutr Hosp. 2015 Feb 26;31 Suppl 3:38–45.

25. De Keyzer W, Huybrechts I, De Vriendt V, Vandevijvere S, Slimani N, Van Oyen H, et al. 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. A novel online Food Recall Checklist for use in an undergraduate student population: a comparison with diet diaries. Nutr J. 2009 Feb 19;8:13.

27. Moshfegh AJ, Rhodes DG, Baer DJ, Murayi T, Clemens JC, Rumpler WV, et al. The US Department of Agriculture Automated Multiple-Pass Method reduces bias in the collection of energy intakes. Am J Clin Nutr. 2008 Jan 8;88(2):324–32.

28. Jacques S, Lemieux S, Lamarche B, Laramée C, Corneau L, Lapointe A, et al. Development of a Web-Based 24-h Dietary Recall for a French-Canadian Population. Nutrients. 2016 Nov 15;8(11):724.

29. Lafrenière J, Lamarche B, Laramée C, Robitaille J, Lemieux S. Validation of a newly automated web-based 24-hour dietary recall using fully controlled feeding studies. BMC Nutr. 2017;3:34.

30. Turgeon O'Brien H, Dufour R. Calcim and magnesium intakes during pregnancy: Association with blood pressure. Current clinical nutrition issues. 1994;74–85.

31. Health Canada. Eating Well with Canada's Food Guide - A Resource for Educators and Communicators [Internet]. Gouvernment of Canada; 2011. Available from: http://www.healthcanada.gc.ca/foodguide

32. Eckel RH, Jakicic JM, Ard JD, de Jesus JM, Houston Miller N, Hubbard VS, et al. 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. 2014 Jul 1;63(25 Pt B):2960–84.

33. Corrao G, Rubbiati L, Bagnardi V, Zambon A, Poikolainen K. Alcohol and coronary heart disease: a meta-analysis. Addict Abingdon Engl. 2000 Oct;95(10):1505–23.

34. Bland JM, Altman DG. Measuring agreement in method comparison studies. Stat Methods Med Res. 1999 Jan 4;8(2):135–60.

35. Masson LF, McNeill G, Tomany JO, Simpson JA, Peace HS, Wei L, et al. Statistical approaches for assessing the relative validity of a food-frequency questionnaire: use of correlation coefficients and the kappa statistic. Public Health Nutr. 2003 May;6(3):313–21.

36. Cade J, Thompson R, Burley V, Warm D. Development, validation and utilisation of food-frequency questionnaires - a review. Public Health Nutr. 2002 Aug;5(4):567–87.

37. Luevano-Contreras C, Durkin T, Pauls M, Chapman-Novakofski K. 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. 2013 Dec;64(8):1030–5.

38. Willett WC, Sampson L, Stampfer MJ, Rosner B, Bain C, Witschi J, et al. Reproducibility and Validity of a Semiquantitative Food Frequency Questionnaire. Am J Epidemiol. 1985 Jan 7;122(1):51–65.

39. Frankenfield D, Roth-Yousey L, Compher C. Comparison of Predictive Equations for Resting Metabolic Rate in Healthy Nonobese and Obese Adults: A Systematic Review. J Am Diet Assoc. 2005 May;105(5):775–89.

40. Heitmann BL, Lissner L, Osler M. Do we eat less fat, or just report so? Int J Obes Relat Metab Disord. 2000 Apr;24(4):435.

41. Labonté M-È, Cyr A, Baril-Gravel L, Royer M-M, Lamarche B. Validity and reproducibility of a web-based, self-administered food frequency questionnaire. Eur J Clin Nutr. 2012 Feb;66(2):166–73.

42. Ambrosini GL, de Klerk NH, O'Sullivan TA, Beilin LJ, Oddy WH. The reliability of a food frequency questionnaire for use among adolescents. Eur J Clin Nutr. 2009 Jun 24;63(10):1251–9.

43. Chiuve SE, Fung TT, Rimm EB, Hu FB, McCullough ML, Wang M, et al. Alternative Dietary Indices Both Strongly Predict Risk of Chronic Disease. J Nutr. 2012 Jan 6;142(6):1009–18.

44. Midanik LT. Validity of Self-reported Alcohol Use: a literature review and assessment. Br J Addict. 1988;83(9):1019–29.

45. National Institutes of Health, National Cancer Institute. Dietary Assessment Primer, 24-hour Dietary Recall [Internet]. [cited 2016 Mar 21]. Available from: http://dietassessmentprimer.cancer.gov/profiles/

**Table 1:** Participants' characteristics (N=107)

| | All participants (N=107) [1] |
|---|---|
| Women (%) | 53.3 |
| Age (years) | 47.4 ± 13.3 |
| Body mass index ($kg/m^2$) | 25.5 ± 4.4 |
| Waist circumference (cm) | 89.1 ± 13.4 |
| Weight stable over the last three months (%) | 57.0 |
| Weight gain over the last three months (%) | 22.4 |
| Weight loss over the last three months (%) | 20.6 |
| Estimated basal metabolic rate – Mifflin (kcal/day) | 1483.8 ± 293.5 |
| Ethnicity – Caucasian (%) | 96.3 |
| Education - High school (%) | 5.6 |
| College (%) | 30.8 |
| University (%) | 63.6 |

[1] Values are presented as means ± standard deviation, or percentage (%)

**Table 2:** Average intakes of energy and nutrients and correlation coefficients between values derived from the web-based self-administered 24-hour dietary recall (R24W) and the food record (FR)

| | **R24W[1] (SD[2])** | **FR[3] (SD)** | **% difference** | **Raw correlation** | **Sex- and energy-adjusted de-attenuated correlation** |
|---|---|---|---|---|---|
| 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* |
| % Carbohydrates | 43.8 (6.9) | 46.7 (7.4) | -6.6* | 0.52* | 0.62* |
| % Fat | 35.8 (6.0) | 35.3 (6.4) | 1.4 | 0.48* | 0.63* |
| % 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 B6 (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 B12 (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* |
| Magnesium (mg) | 463.7 (149.5) | 461.1 (177.8) | 0.6 | 0.52* | 0.65* |
| Zinc (mg) | 14.1 (4.8) | 12.8 (3.9) | 9.2* | 0.38* | 0.02 |
| Iron (mg) | 17.2 (4.9) | 16.5 (4.9) | 4.1 | 0.34* | 0.46* |
| Calcium (mg) | 1281.4 (450.3) | 1117.3 (396.7) | 12.8* | 0.53* | 0.50* |
| Potassium (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* |
| Saturated fatty acids (g) | 35.4 (17.8) | 30.7 (15.1) | 13.3* | 0.58* | 0.47* |
| Sodium (mg) | 3455.4 (1127.0) | 3154.9 (1110.0) | 8.7* | 0.55* | 0.36* |
| Average (SD) | | | 1.7 (16.9) | 0.48 (0.09) | 0.52 (0.15) |

* Student’s t-test and Pearson correlation with a P-value < 0.05
[1] R24W= Web-based self-administered 24-hour dietary recall
[2] SD= Standard deviation
[3] FR= Food record

**Table 3:** Cross-classification of energy and nutrient intakes into quartiles of the distribution using either the web-based self-administered 24-hour dietary recall or the food record

| | % | | | | |
|---|---|---|---|---|---|
| | **Same quartile** | **Adjacent quartiles** | **± 1 Quartile apart** | **Misclassification (quartile 1 vs. 4)** | **Kappa score** |
| 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 |
| % Carbohydrates | 34.6 | 42.1 | 76.6 | 4.7 | 0.25 |
| % Fat | 35.5 | 42.1 | 77.6 | 2.8 | 0.28 |
| % 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 B6 (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 B12 (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 |
| Magnesium (mg) | 42.1 | 34.6 | 76.6 | 4.7 | 0.31 |
| Zinc (mg) | 31.8 | 40.2 | 72.0 | 2.8 | 0.20 |
| Iron (mg) | 36.4 | 42.1 | 78.5 | 4.7 | 0.28 |
| Calcium (mg) | 39.3 | 44.9 | 84.1 | 1.9 | 0.37 |
| Potassium (mg) | 36.4 | 46.7 | 83.2 | 1.9 | 0.34 |
| Alcohol (g) | 50.5 | 33.6 | 84.1 | 1.9 | 0.47 |
| Saturated fatty acids (g) | 37.4 | 45.8 | 83.2 | 2.8 | 0.34 |
| Sodium (mg) | 46.7 | 33.6 | 80.4 | 1.9 | 0.40 |
| Average (SD[1]) | 40.0 (5.7) | 40.0 (5.0) | 80.0 (4.3) | 3.6 (1.7) | 0.33 (0.1) |

[1] SD= Standard deviation

**Table 4:** Statistical test outcomes and proportion of poor outcomes for the relative validity of the web-based self-administered 24-hour dietary recall (R24W)

| | 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 | Kappa Statistic | % difference | Student's t-test | Bland Altman (slope of the regression) | Proportion of poor outcomes (/7) |
| Criteria for good outcome (G) | ≥0.50 | ≥ 50% in the same quartile; <10% in opposite quartile | ≥0.61 | 0-10.9% | P>0.05 | P>0.05 | |
| Criteria for acceptable outcome (A) | 0.20-0.49 | | 0.20-0.60 | 11.0-20% | | | |
| Criteria for poor outcome (P) | <0.20 | <50% in the same quartile; ≥10% in opposite quartile | <0.20 | >20% | P≤0.05 | P≤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 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| % Carbohydrates | G | P-G | A | G | P | G | 2 |
| % Fat | G | P-G | A | G | G | G | 1 |
| % 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 B6 (mg) | A | P-G | A | G | P | G | 2 |
| Folic acid (µg) | A | P-G | A | G | G | G | 1 |
| Vitamin B12 (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 |
| Magnesium (mg) | G | P-G | A | G | G | P | 2 |
| Zinc (mg) | P | P-G | A | G | P | P | 4 |
| Iron (mg) | A | P-G | A | G | G | G | 1 |
| Calcium (mg) | G | P-G | A | A | P | G | 2 |
| Potassium (mg) | G | P-G | A | G | G | G | 1 |
| Alcohol (g) | G | G-G | A | P | P | P | 3 |
| Saturated fatty acids (g) | A | P-G | A | G | P | G | 2 |
| Sodium (mg) | A | P-G | A | G | P | G | 2 |

G=Good outcome; A=Acceptable outcome; P=Poor outcome

**Table 5:** Proportion of under-, adequate and over-reporters assessed by the ratio between reported energy intake and basal metabolic rate as determined with the web-based self-administered 24-hour dietary recall (R24W) and the food record (FR).

| | **Total (107)** | | **Men (50)** | | **Women (57)** | |
|---|---|---|---|---|---|---|
| | **R24W**[1] | **FR**[2] | **R24W** | **FR** | **R24W** | **FR** |
| **Under-reporters (ratio[3]<1.35)** | 15.0% | 23.4% | 20.0% | 24.0% | 10.5% | 22.8% |
| **Adequate reporters (ratio 1.35-2.5)** | 80.4% | 74.8% | 72.0% | 72.0% | 87.7% | 77.2% |
| **Over-reporters (ratio>2.5)** | 4.7% | 1.9% | 8.0% | 4.0% | 1.7% | 0.0% |

[1] R24W=Web-based self-administered 24-hour dietary recall
[2] FR= Food record
[3] Ratio=reported energy intake/basal metabolic rate