Repeatability of Food Frequency Assessment Tools in relation to the number of items and response categories included

Running head: Repeatability of nutritional assessment tools

Vassiliki Bountziouka, Eirini Bathrellou, Itziar Zazpe, Leyre Ezquer, Miguel-Angel Martínez-González, Demosthenes B. Panagiotakos

Vassiliki Bountziouka, Eirini Bathrellou, Demosthenes B. Panagiotakos, are affiliated with Harokopio University, Athens, Greece
Itziar Zazpe, Leyre Ezquer and Miguel-Angel Martínez-González are affiliated with Navarra University, Navarra, Spain

Address for correspondence
Dr. Demosthenes B Panagiotakos
46 Paleon Polemiston St., 166 74, Glyfada, Attica, Greece,
E-mail: dbpanag@hua.gr
Tel.: +30 210 9603116; Fax: +30 210 9600719

Abstract

Background: Accuracy of a measurement is a cornerstone in research in order to make robust conclusions about the research hypothesis. **Objective:** To examine whether the number of items (questions) and the number of responses of consumption included in nutritional assessment tools influence their repeatability. **Methods:** During 2009, 400 participants (250 from Greece, 37±13 yrs, 34% males and 150 participants from Spain, 39±17 yrs, 41% males) completed a diet index with 11-items and binary (yes/no) responses, a diet-index with 11-items and 6-scale responses, a 36-item and a 76-item food frequency questionnaire (FFQ) with 6-scale responses. Participants completed these tools, twice, within 15-days period. Spearman-Brown (rsb), Kendall's tau coefficients and the Bland-Altman method were applied to answer the research hypothesis. **Results:** The highest repeatability coefficient was observed for the 11-items with binary responses index (rsb=0.948, p<0.001), followed by the 11-items with 6-scale responses index (rsb=0.943, p<0.001), the 36-item (rsb=0.936, p<0.001) and the 76-item FFQs (rsb=0.878, p<0.001). Statistical comparisons revealed no significant differences between repeatability coefficients of the first three tools (p>0.23); whereas the aforementioned tools had significantly higher repeatability coefficients as compared with the 76-item FFQ (p=0.002). Sub-group analyses by gender, education, smoking and clinical status, confirmed the aforementioned results. **Conclusion:** Repeatability has been revealed for all food frequency assessment tools used, irrespective of the number of items or the number of responses included.

Key words: assessment tools; accuracy; repeatability; methodology.

Introduction

An accurate dietary assessment is of major importance in research in order to make robust conclusions, and especially when the role of diet on the development of various diseases is examined [1]. Common practice in the majority of observational and clinical studies is to ask participants to report their dietary habits with the use of questionnaires or diaries [2]. Furthermore, during the past years several diet indices (i.e., modules or brief FFQs) have also been developed and proposed in order to measure quality of diet through the level of adherence to specific dietary patterns (e.g., Mediterranean Diet Scale, Mediterranean Diet Score, MedDietScore) or dietary guidelines suggested by various Organizations (e.g., Healthy Eating Index, Diet Quality Index, etc) [3]. These indices are developed using a small number of items (questions), usually between 9-15 and a variety of consumption responses [i.e., binary (yes/no), or multinomial (never/rare/monthly/weekly/daily)] [4].

As already mentioned, a major challenge the scientists face in nutrition assessment is the reliability of the dietary information retrieved through the various nutrition assessment tools (e.g., dietary recalls, food records, diet history, and food frequency questionnaire) [5-7]. A crucial component of the reliability in the retrieved information is the repeatability of the nutrition assessment tool, i.e., the variation in consecutive measurements taken by a tool of the same individual and under the same conditions. Good repeatability suggests that the tool has been appropriately designed and does not allow random errors in the information collected. A methodological issue that may affect the level of repeatability of a nutrition assessment tool is the number of items (questions) included. Long FFQs capture more information regarding nutrient intake, but they may also be exhausting in face-to-face interviews having an impact in the repeatability, and hence, at least in part, the reliability of the information retrieved. In contrast, short FFQs may be easily and timely accessible, but they are lacking of nutrient information. Thus, the question regarding the optimal number of items used to develop an accurate and informative nutritional assessment tool is still open [8]. In addition, it remains unsolved whether the number of consumption responses used (i.e., the coding used to measure the frequency of consumption, e.g., “yes/no” or more detailed, “never/rare/1-2 times per month/.../daily”) may influence the level of repeatability of a tool [9, 10].

Therefore, the aim of this work was to evaluate to which extent the number of items included in a nutrition assessment tool (i.e., FFQ or diet index), as well as the number of consumption responses of each item, influences the repeatability of the measurement. To test the research hypothesis four tools were developed and were used in a sample of 400 subjects from Greece and Spain. Two of them had the same number of questions but different response classes (i.e., a diet index with 11-items and binary responses, a diet-index with 11-items and 6-scale responses) and the other two had different number of questions but same response classes (i.e., a 36-item and a 76-item food frequency questionnaire (FFQ) with 6-scale responses).

Methods

Participants

From March 2009 till February 2010, 400 participants, 250 from Greece (37±13 yrs, 34% males) and 150 from Spain (39±17 yrs, 41% males), were enrolled in the study on a voluntary basis (participation rate 85%) to facilitate the generalization of the results in other populations as well. The sample size was considered adequate in achieving statistical power equal to 99% for the evaluation of two-sided mean differences equal to 1.0±0.5 times per week for the consumption of various foods at 0.05 type I error based on the 76-item FFQ. The data were confidential and the study followed the ethical considerations provided by the World Medical Association (52nd WMA General Assembly, Edinburgh, Scotland, October 2000). Moreover, the Ethics Committee of the Harokopio University approved the design, procedures and aims of the study (GA 23/14.05.2009). All participants were informed about the aims and purpose of the study and gave their consent.

Nutrition Assessment Tools

Participants were asked to complete twice, during the same period, through face-to-face interview by trained dieticians, a series of nutrition assessment tools within 15-days interval. According to Streiner and Norman, this time interval can be considered as the shorter time that a person cannot recall previous responses and as such was selected to avoid potential alterations in dietary intake (e.g., due to fasting period) [11].

Participants completed: (a) a 11-item diet index with binary coding for consumption responses (named here as “Diet Index”), (b) a 11-item with the with 6-scale coding for consumption responses diet index (the MedDietScore, theoretical range 0-55) [12], as well as: (c) a 36-item and (d) a 76-item semi-quantitative FFQs, that has been previously validated [13] (Appendix Table). All the questionnaires used for the Greek sample were literally translated to Spanish by a group of experts. The reference period for the collection of food data using all tools was the past one month of intake. The “Diet Index” included 11 food questions regarding the recommended by the Hellenic Ministry of Health and Welfare [14] consumption of: non-refined cereals (if 25-30 portions/ week), potatoes (1-3 portions/ week), fruit and juice (≥10 portions/ week), vegetables (≥20 portions/ week), pulses (1-3 portions/ week), fish (2-4 portions/ week), red meat and meat products (≤1 portions/ week), poultry (1-2 portions/ week), full fat dairy (6-8 portions/ week), daily use of olive oil, and alcohol drinks (1-2 glasses/ day). Score 1 was given if the consumption was according to the recommendations, otherwise score 0 (theoretical range 0-11). In order to test the hypothesis whether the number of item’s response categories is related to the repeatability of the tool, under the condition of the same number of items used, the MedDietScore included the same 11 questions as the “Diet Index” with a wider range of possible responses in each question. Highest values of this diet score indicates greater adherence to the Mediterranean diet, whereas the MedDietScore has also been previously validated. Details regarding the scoring system of the MedDietScore can be found elsewhere [12]. All diet indices aimed to evaluate adherence to the Mediterranean Diet [15].

As regards the 36-item FFQ, information on «dairy» products included full fat and low fat milk/ yoghurt, all kind of cheese and egg; «starchy» food group included questions regarding the consumption of bread/ crisps, breakfast cereals, rice/ pasta/ other kind of cereals, bakery products, homemade or ready to eat pies and potatoes; under «meat» category five questions were included regarding the frequency of beef meat, pork meat, chicken, lamb and delicatessens consumption. Additionally, information on fish and seafood consumption was also included. Foods of plant origin included pulses/ legumes, all kind of vegetables as a salad dish and as main dish, fresh/ dried fruit and nuts. The «sweet» group included the following food items/ groups: chocolate/ biscuits, Greek sweets and ice cream/ cream/ rice pudding. Information regarding beverages consumption referred to consumption of all kinds of alcohol, sodas, fruit juice, coffee and tea intake. The 76-item FFQ included the aforementioned categories as regards the food items in more detail. Both FFQs had 6-scale response categories (i.e., 1: rarely/ never, 2: 1-3 times/ month, 3: 1-2 times/ week, 4: 3-6 times/ week, 5: 1 time/ day, 6: ≥2 times/ day). More information about the FFQ used may be found elsewhere [13]. Moreover, the MedDietScore was also indirectly calculated from the 76-item FFQ to test the agreement between the directly and indirectly estimates of a diet index.

Other Measurements
Information was also collected regarding basic demographic characteristics, such as age, gender, educational status according to years of schooling, current smoking status (yes/no), as well as a short medical history (i.e., presence of hypertension, dyslipidaemia, diabetes, cardiovascular disease, renal failure and cancer). In addition participants' body height (in meters) and weight (in kilograms) was recorded. These factors were taken into account to perform sub-group (sensitivity) analyses to further evaluate the research hypothesis.

Statistical Analysis
The method suggested by Bland and Altman (B&A) was used to assess the repeatability between the two administrations of the dietary tools (FFQs and indices). In particular, the differences in the recordings (as transferred into daily intake) of both administrations and as the average of the two were calculated according to the protocol [16]. Furthermore, the Spearman correlation coefficient (Spearman's rho) between the difference and the average for each item was calculated to assess potential bias between the rankings of the difference and the average. According to this method, the narrower the limits of agreement (i.e., $mean_{(difference)} \pm 1.96 * standard\ deviation_{(difference)}$) the better the repeatability, while the correlation coefficient should be close to zero indicating lack of bias between the test-retest assessment tool. In addition, the degree of repeatability of the nutrition assessment tools (i.e., the 36- and the 76-item FFQ, and the dietary indices as well) was also evaluated by the Spearman-Brown (r_{sb}) coefficient (values close to +1 suggest good repeatability) and Kendall's-tau *b* coefficient (values close to +1 indicating good agreement between the two administrations) that used case-by-case data. Comparisons between the repeatability coefficients were made using the Fisher transformation and the Z-test. The normality of data distribution was tested using P-P plots and group comparisons were performed using the paired Student's t-test. Descriptive characteristics of the participants are presented as mean±SD and absolute (n) and relative (%) frequencies. Sub-group analyses were performed by gender (males vs. females), education status (basic, i.e., <12 years of school, vs. higher), current smoking (yes vs. no) and clinical status [no risk factor (i.e., obesity, history of hypertension, diabetes mellitus, hypercholesterolemia, cardiovascular disease, renal failure, cancer) vs. at least one factor]. SPSS version 18 (Statistical Package for Social Sciences, SPSS Inc, Chicago, IL, U.S.A.) software was used for all the statistical calculations.

Results

Participants' characteristics are presented In *Table 1*.
(Table 1)

In general, all diet indices and FFQs showed very good repeatability, as all coefficients were higher than 0.80. The highest repeatability coefficient was observed for the 11-items index with binary responses, followed by the 11-items with 6-scale responses index, the 36-item, and finally, the 76-item FFQs (*Table 2*). In addition, the mean total scores between the test-retest administrations of each diet index were similar regarding the nutritional meaning of the mean score values (Diet Index: 5.06±1.78 vs. 5.69±1.83, p=0.10, and MedDietScore: 24.7±4.41 vs. 24.9±4.57, p=0.03), also suggesting good repeatability. The Kendall's-tau coefficients were high (i.e., >0.50) and thus, confirmed the aforementioned results. In particular, Kendall's-tau coefficients ranged from 0.71 for potatoes to 0.89 for pulses, with respect to the Diet Index; from 0.38 for potatoes to 0.86 for red meat, with respect to the MedDietScore; from 0.68 for rice/ pasta to 0.90 for coffee, with respect to the 36-item FFQ; and from 0.40 for starchy vegetables (i.e., petit pois (peas), green beans, okra) to 0.74 for cereals and coffee, with respect to the76-item FFQ (*Figure 1*). The B&A method showed close to zero mean differences between the test-retest administrations (i.e., suggesting lack of bias) and relatively narrow limits of agreement in all tools tested (*Table 2*). However, statistical comparisons revealed that the 11-items index with binary responses, the 11-items with 6-scale responses index and the 36-item FFQ had significantly higher repeatability coefficients as compared with the 76-item FFQ (p<0.002), while no significant differences between the repeatability coefficients of the two 11-items indices and the 36-item FFQ were observed (p>0.23), indicating that as the number of food items increased the repeatability of the tool tend to decrease. At this point it should be noted that the differences in absolute values is meaningless in terms of nutritional information.

The MedDietScore was, afterwards, indirectly calculated through the 76-item FFQ in order to test its repeatability when calculated through this procedure (which is common in research). Both indirectly and directly MedDietScore were found repeatable (27.7±3.37 vs. 27.5±3.72, p=0.32; 24.7±4.41 vs. 24.9±4.57, p=0.03, respectively); however, the repeatability coefficient was higher for the directly as compared with the indirectly calculated MedDietScore (p<0.001). Moreover, it was observed that the total value of the indirectly calculated MedDietScore was higher as compared with the directly calculated from the participants in each administration of the tool (both p's <0.001), which may not reflect true adherence with the guidelines of this specific pattern.

(Table 2)

(Figure 1)

Sub-group Analyses

Sub-group analyses were performed to evaluate the repeatability of the dietary indices by gender, education status, smoking habits and medical history of the participants (*Table 3*). Almost all three diet indices were repeatable in all sub-groups tested. However, lack of repeatability was noticed for the 11-item index with binary responses as regards the participants who had at least one clinical factor in their medical history and for the 11-item index with 6-scale responses as regards female gender and smokers. Despite the significance of the aforementioned results it should be noted that the observed mean differences between the test and retest administrations were meaningless in terms of their nutritional information. For example, the difference of 0.6 units in the two administrations of the MedDietScore (i.e., the 11-item index with 6-scale responses) as reported by females, although statistical significant does not indicate differences in mean frequent consumption of food groups that consist of the MedDietScore. Finally, no differences were observed in terms of repeatability level between the Spanish and the Greek groups (data not shown here).

(Table 3)

Discussion

In this methodological work it was revealed that the level of repeatability of food frequency assessment tools, a crucial part of their reliability, is not influenced by the number of food items included in each dietary tool, as well as the number of consumption responses used for each food item. From a statistical point of view the lowest repeatability level was observed for the more detailed FFQ (i.e., with 76-items), and as the number of food items increased the repeatability of the tool tended to decrease, but the difference between the coefficients, although significant, was of limited nutritional interest. It is also important to note that all dietary tools used here have been previously validated. Sensitivity analyses confirmed these results in almost all sub-groups tested; however, special attention should be given in order to accurately collect dietary information on, the gender of the responders, the education status, smoking habits and medical history. The presented findings may be considered of major importance in both research and public health, since investigators and clinicians may use comprehensive FFQs or diet indices with a variety number of consumption responses, in order to retrieve consistent information. The use of two samples from different cultures, i.e., Greek and Spanish, enhance the importance of the results. Nevertheless, to the best of our knowledge, information regarding the repeatability of a nutritional assessment tool and its' relation to the length and the depth of the tool used is lacking, thus, further validation studies in other cultures are required to confirm or refute the aforementioned finding.

The level of repeatability of the dietary assessment tools used in this work was very good, as all the statistical criteria applied had high values. In general, the level of repeatability observed in various other similar studies varies between 0.6 and 0.9 [2]; however, direct comparisons are hard to be made because of the different statistical criteria used. An important issue that was particularly tested here was the repeatability of FFQs and diet indices, in terms of the number of response categories used for each food component, as well as the way of the calculation of the index (direct or indirect through an FFQ). The latter research question is of importance because although the use of FFQs is a common practice in nutritional epidemiology, during the past years several composite diet indices have also been developed to monitor population's adherence to dietary guidelines, as well as screening tools regarding the quality of diet [10, 17]. The use of composite indices is not only attractive due to their practical use, but also mandatory in order to address inferential problems derived in data analysis, mainly caused by the synergistic effects between several inherent characteristics of the diet, which express different dimensions of an attribute (e.g., increased meat and reduced fruits and vegetables consumption) [18, 19]. However, a number of methodological issues have been raised regarding the development of a composite index (e.g., dietary score) [18]. For example, the number of response categories that should be used has not been fully understood and clarified. It has been suggested that in terms of consistency and stability of the responses over short periods, a dichotomous (i.e., yes/no) index may lead in more robust, repeatable results, and thus, it's' use tends to be preferable; whereas, others have suggested that large scale partitioning may lead to more sensitive, in terms of health outcomes, indices [19, 20, 21]. The presented findings showed that the level of repeatability of indices was similar, irrespectively of the number of responses used to evaluate frequency of consumption. The diet index with the binary responses showed slightly higher repeatability as compared with the 6-scale directly and indirectly calculated MedDietScore (*Table 2*). Therefore, health practitioners can use more detailed responses than a simple "yes/no" scale in the development of an index, without losses in the accuracy of the information retrieved (as shown here), but gaining in diagnostic accuracy, as recently reported [19]. In line with the presented findings is another study in elderly Australians where the 15-item FFQ was comparable with the 35-item FFQ in terms of repeatability; and the authors concluded that "*the short-item FFQ was the preferred tool to use in clinical and research settings, especially for older people, as it requires less time and effort to complete*" [22].

Diet indices can be also calculated indirectly through a FFQ that has previously been administrated. In fact, this approach is often used, since researchers usually apply a FFQ and then calculate the index using the information retrieved through the questionnaire. As far as it concerns these two approaches for calculating a diet index (i.e., directly and indirectly), the presented results suggest that the repeatability coefficients were higher for the directly calculated MedDietScore (*Table 2*). Although the difference in the coefficients was small (i.e., 0.943 vs. 0.870), this may suggests that a direct calculation seems preferable to achieve repeatability. Moreover, differences in participants' scores showed lack of bias for the test-retest administration, suggesting that the repeatability of the diet indices was achieved for those that scored low, as well for those that scored higher values. However, it should be underlined that the indirectly calculated MedDietScore was higher than the directly (*Table 3*); a fact that may lead to the conclusion that the calculation of a diet quality index through a FFQ may result to an overestimation of the quality of diet (if the index was designed for). An explanation for this could be the suggestion by other researchers that longer FFQs may overestimate consumption of foods, especially fruits and vegetables [23].

Limitations

Only two regions were used, Spain and Greece, therefore cultural and behavioural differences may occur in other places of the world were not considered in this work, and thus, the generalization of the findings to other populations should be made with conscious. In addition, although repeatability of all dietary tools was revealed, their validity was not tested. As regards the tools used to evaluate the research hypothesis, the lack of a much longer FFQ (i.e., 150-items) limits the observed inverse trend between repeatability level and number of items included in the nutrition assessment tool. Finally, although specific attention was paid for the time interval within the two administrations to avoid correlated errors, the results should be also confirmed considering a longer lapse time.

Conclusion

In this work the repeatability of various nutritional assessment tools was evaluated in relation to the number of items and the number of consumption responses used. The data analysis revealed that the level of repeatability was irrespective of the number of items and consumption responses used in the development of a FFQ or an index. Nevertheless, it should be also mentioned that the highest repeatability coefficients were noticed for small length questionnaires. Thus, investigators should decide whether they need to capture all the required nutritional information regarding dietary habits or to assess overall dietary habits and, therefore, use long or short length questionnaires, respectively, without endanger the consistency of the information retrieved. Understanding the purpose of the assessment will ensure the appropriate method of evaluation to be used.

Acknowledgement

The authors are, particularly, grateful to the men and women from the residence of Athens, Greece and Navarra, Spain, who participated in and collaborated on this research. They also wish to express their gratitude to E. Pagourtzi, E. Serpanou and K. Tsoutsoulopoulou, field investigators from the Harokopio University, as well as Dr E. Polychronopoulos and Dr T. Constantinides, for their substantial assistance in the enrolment of the participants.

The authors state that there are no conflicts of interest.

VB conducted the research, analyzed the data and wrote the paper; EB, IZ & LE contributed to acquisition of data and critically revise the draft for important intellectual content; MAMG study oversight, critically revise the draft for important intellectual content; DBP, had project conception, development of overall research plan, study oversight, primary responsibility, wrote and final approval for final content of the paper.

References

1. National Research Council-Committee on Diet and Health. Diet and health implications for reducing chronic disease risk. Washington, DC: National Academy Press; 1989: 651-664.
2. Willett WC. Food Frequency Methods. In: Willett WC, ed. Nutritional Epidemiology. 2nd ed. New York: Oxford University Press;1998.
3. Bach A, Serra-Majem L, Carrasco JL, Roman B, Ngo J, Bertomeu I, Obrador B. The use of indexes evaluating adherence to the Mediterranean Diet in epidemiological studies: a review. Publ Health Nutr 2006;9:132-46.
4. Kourlaba G, Panagiotakos D. Dietary quality indices and human health: a review. Maturitas 2009;62:1-8.
5. Beaton GH, Burema J, Ritenbaugh C. Errors in the interpretation of dietary assessments. Am J Clin Nutr 1997;65:1100S-1107S.
6. Golafshani N. Understanding Reliability and Validity in Qualitative Research. The Qualitative Report 2003;8:597-607.
7. Kubena KS. Accuracy in dietary assessment: On the road to good science. J Am Diet Assoc 2000;100:775-776.
8. Grant A, DeHoog S. Nutrition Assessment and Support. 3rd ed. Seattle: WA; 1999.
9. O'Leary CJ, Jones PW. The influence of decisions made by developers on health status questionnaire content. Qual Life Res 1998;7:545–550.
10. Panagiotakos DB. Health measurement scales: methodological issues. Open Cardiovasc Med J 2009;3:160-165.
11. Streiner DL, Norman GR. Health Measurement Scales: A Practical Guide to Their Development and Use. Oxford Medical Publications;1995.
12. Panagiotakos DB, Pitsavos C, Arvaniti F, Stefanadis C. Adherence to the Mediterranean food pattern predicts the prevalence of hypertension, hypercholesterolemia, diabetes and obesity, among healthy adults; the accuracy of the MedDietScore©. Prev Med 2007;44:335-340.
13. Bountziouka V, Bathrellou E, Giotopoulou A, Katsagoni C, Bonou M, Vallianou N, Barbetseas J, Avgerinos PC, Panagiotakos DB. Development, repeatability and validity regarding energy and macronutrients intake of a semi-quantitative Food Frequency Questionnaire. Nutr Metabolism Cardiovasc Dis 2010;doi:10.1016/j.numecd.2010.10.015.
14. Ministry of Health and Welfare, Supreme Scientific Health Council. Dietary guidelines for adults in Greece, 1999.
15. Trichopoulou A, Costacou T, Bamia C, Trichopoulos D. Adherence to the Mediterranean Diet and survival in a Greek population. N Engl J Med 2004;348:2599-2608.
16. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;1:307-310.
17. Kant AK, Thompson FE. Measures of overall diet quality from a food frequency questionnaire:National Health Interview Survey, 1992. Nutr Res 1997;17:1443-1456.

18. Hu FB. Dietary pattern analysis:a new direction in nutritional epidemiology. Curr Opin Lipidiol 2002;13:3-9.
19. Kourlaba G, Panagiotakos DB. The diagnostic accuracy of a composite index increases as the number of partitions of the components increases and when specific weights are assigned to each component. J Appl Stat 2010;37:537-554.
20. Kourlaba G, Panagiotakos DB. The number of index components affects the diagnostic accuracy of a diet quality index: The role of intra- and inter-correlation structure of the components. Ann Epidemiol 2009;19:692-700.
21. Oshaug A. Dietary Assessment Methods. In: FAO eds. Methods to monitor the human right to adequate food - Volume II. Rome; 2008:144-149.
22. Miller M, Chi Yeo Y, Jiun Khor M, Clover E, Magarey A. Repeatability of a Short Food Frequency Questionnaire to Assess Calcium Intake in Older Australians. J Aging Res. 2010; 2010: 905056.
23. Krebs-Smith SM, Heimendinger J, Subar AF, Patterson BH, Pivonka E. Estimating fruit and vegetable intake using food frequency questionnaires: a comparison of instruments. Am J Clin Nutr 1994;59:283S.

Table 1: Socio-demographic, lifestyle and clinical characteristics of the participants in the study (n=400).

	Greece	**Spain**
N	250	150
Age, (yrs)	37±13	39±17
Males, N (%)	85 (34)	61(41)
Highest educational status (>12yrs), N (%)	163 (65)	59 (39)
Smoking habits (current), N (%)	104 (42)	43 (29)
Prevalence of obesity, N (%)	27 (11)	5 (3.3)
Prevalence of hypertension, N (%)	16 (6.4)	14 (9.3)
Prevalence of dyslipidemia, N (%)	22 (8.8)	11 (7.3)
Prevalence of diabetes, N (%)	6 (2.4)	6 (4.0)
Prevalence of renal failure, N (%)	0 (0.0)	2 (1.3)
History of chronic diseases (i.e., cardiovascular, cancer), N (%)	6 (2.4)	2 (1.3)

Table 2: Results of repeatability between dietary assessment tools, based on Spearman-Brown (r_{sb}) coefficients and Bland & Altman Mean Difference and Limits of Agreement, in the whole study's participants (n=400).

Assessment Tool	r_{sb}	Mean Difference	Limits of Agreement
"Diet Index", 11-item, binary responses	0.948*	-0.08	(-1.66, 1.49)
MedDietScore†, 11-item, 6-scale responses	0.943*	-0.41	(-5.75, 4.94)
MedDietScore‡, 11-item, 6-scale responses	0.870*	0.16	(-4.79, 5.11)
FFQ, *36-item,* 6-scale responses	0.936*	N/A	N/A
FFQ, *76-item,* 6-scale responses	0.878*	N/A	N/A

r_{sb} = Spearman – Brown correlation coefficient. FFQ= Food Frequency Questionnaire
N/A = not applicable. Bland and Altman mean difference and limits of agreement cannot be calculated for the questionnaires.
† Directly calculated.
‡ Indirectly calculated through the 76-item FFQ (calculated).
* $p<0.001$ (H_o: $r_{sb} = 0$)

Table 3: Results regarding the repeatability of the diet indices, for the whole sample (*n*=400) and by selected sub-groups (Mean value ± Standard Deviation, SD).

	Diet Index (0-11)			**MedDietScore (0-55)**†			**MedDietScore (0-55)**‡		
	Test mean±SD	**Retest mean±SD**	***P***	**Test mean±SD**	**Retest mean±SD**	***P***	**Test mean±SD**	**Retest mean±SD**	***P***
Overall	5.06 ± 1.78	5.69 ± 1.83	0.10	24.70 ± 4.41	24.90 ± 4.57	0.03	27.70 ± 3.73	27.54 ± 3.72	0.32
Gender									
Males	5.47±1.78	5.59±1.87	0.16	24.8±4.64	24.7±4.50	0.77	27.31±3.63	27.08±3.58	0.43
Females	5.67±1.78	5.74±1.81	0.30	24.6±4.31	25.2±4.61	0.001	27.90±3.78	27.77±3.77	0.51
Educational status									
Basic (≤12 yrs)	5.73±1.80	5.79±2.01	0.49	23.2±4.17	23.4±4.05	0.55	27.45±3.96	27.14±3.72	0.36
Higher (>12 yrs)	5.50±1.75	5.55±1.67	0.42	25.2±4.38	25.6±4.57	0.06	27.58±3.67	27.60±3.78	0.92
Smoking									
No	5.75±1.76	5.85±1.74	0.15	25.2±4.34	25.3±4.41	0.38	27.66±3.57	27.38±3.49	0.17
Yes	5.40±1.80	5.46±1.93	0.43	24.0±4.44	24.5±4.75	0.02	27.75±3.98	27.75±4.03	0.99
Clinical status									
No factor	5.68±1.79	5.74±1.83	0.29	24.8±4.30	25.1±4.39	0.09	27.69±3.87	27.60±3.85	0.65
At least one factor	5.36±1.73	5.50±1.83	0.02	24.1±4.78	24.6±5.15	0.13	27.73±3.26	27.30±3.24	0.19

Paired comparisons of the test-retest administrations were performed with Student's t-test, after controlling for the normality of the distribution using P-P plots.
† Directly calculated.
‡ Indirectly calculated through the 76-item FFQ.

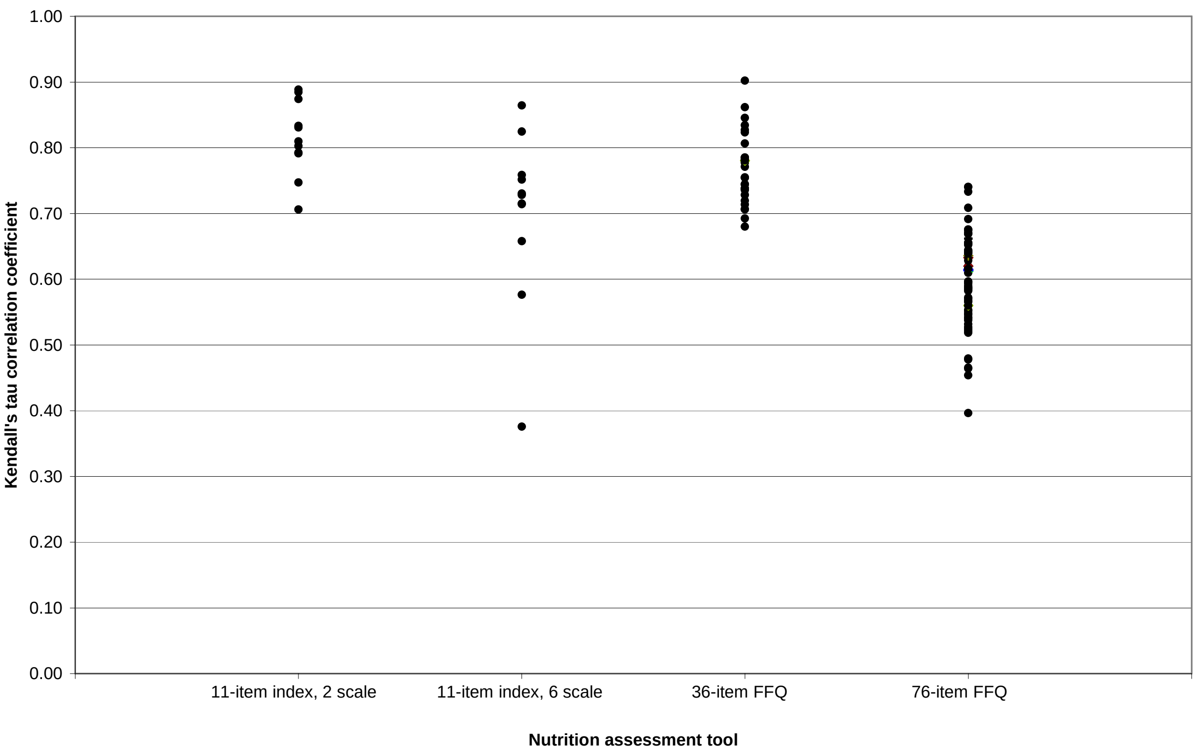

Figure 1: Distribution of Kendall's tau-b correlation coefficients (theoretical range, -1 to +1) between the two administrations of each nutrition assessment tool as regards the items (questions) included (dots represent each item).

Appendix Table. Food items (i.e., questions) that were used for the development of the nutritional assessment tools.

Nutritional Assessment Tools		
Diet Index & MedDietScore	**36 -item FFQ**	**76-item FFQ**
Full-fat dairy products	Full-fat milk, yoghurt	Full-fat milk, yoghurt
Non-refined cereals	Low-fat milk, yoghurt	Low-fat milk, yoghurt
Potatoes	Cheese	Yellow cheese, cream cheese
Red meat and products	Egg (boiled, fried, omelet)	Feta cheese, anthotiro cheese
Poultry	Bread, crispies	Low fat cheese
Fish	Bakery	Egg (boiled, fried, omelet)
Legumes	Pies (e.g. cheese pies, spinach pies)	White bread
Vegetables	Cereals	Whole meal bread, rusks
Fruit	Rice, pasta	Pita bread, burger-bread
Alcohol	Potatoes (boiled, mashed, baked, fried)	Crisp breads
Olive oil	Beef	Breakfast cereal, cereal bars
	Pork	White rice (1 cup)
	Lamp	Brown rice (1 cup)
	Meat products	Pasta, pearl barley
	Poultry	Whole meal pasta
	Fish	Potatoes (boiled, mashed, baked)
	Seafood	Fried potatoes
	Pulses	Home made pies (e.g. cheese-pie, spinach-pie)
	Vegetables	Pies (e.g. cheese pies, spinach pies)
	Salads	Toasted sandwich, sandwich
	Fruit fresh, dried fruit	Veal
	Nuts	Burger/ meat balls/ minced-collops
	Sweets made in tray	Pork (steak,filet,souvlaki)
	Chocolate	Lamb/ goat/ game/ lambchops
	Ice cream, rice pudding	Cold sliced meats
	Fruit juice	Sausage/ bacon
	Soft drinks	Light/ no fat cold sliced meats
	Alcohol	Chicken, turkey
	Coffee	Small fish
	Tea	Large fish
		Sea-food (octopus, sleeve-fish, prawns)
		Lentils, beans, chickpeas
		Spinach-rice/ cabbage-rice
		Pastitsio/ mousakas/ papoutsakia
		Petit pois (peas), green beans, okra, artichoke
		Tomato, cucumber, carrot, pepper
		Lettuce, cabbage, spinach, rocket
		Broccoli, cauliflower, courgette
		Greens, celery, spinach
		Orange

Apple, pear
Other winter-fruits
banana
Other summer-fruits
Fruit juice (1 glass)
Dried fruits (1/4 cup)
Nuts(1 little cup)
Sweets made in tray
Sweet preserves, stewed fruit, fruit-jelly
Gateau, tart
Croissant, gofer, cake, biscuits
Chocolate
Ice-cream, milkshake, cream, rice pudding
Chips, pop-corn
Honey, marmelade, sugar
Olives
Wine
Beer
Other alcohol drinks
Soft drinks
Light soft drinks
Coffee
Tea, herbal tea
Mayonnaise, sauce
Light mayonnaise, light sauce
Olive oil
Seed oil
Margarine
Butter

10