A systematic review of brief dietary questionnaires suitable for clinical use in the prevention and management of obesity, cardiovascular disease and type 2 diabetes

England, CY; Andrews, RC; Jago, R; Thompson, JL

DOI:
10.1038/ejcn.2015.6

Document Version
Peer reviewed version

Citation for published version (Harvard):
England, CY, Andrews, RC, Jago, R & Thompson, JL 2015, 'A systematic review of brief dietary questionnaires suitable for clinical use in the prevention and management of obesity, cardiovascular disease and type 2 diabetes', *European Journal of Clinical Nutrition*, vol. 69, pp. 977–1003. https://doi.org/10.1038/ejcn.2015.6

Table 2: Summary of reliability and relative validity

Tool	Test retest reliability					Validity				
	Retest time	**Variables**	**Test**	**Results**	**Internal reliability**	**Reference measure**	**Time between tests**	**Variables**	**Test**	**Results**
Healthy Eating										
Australian Diet Quality Tool (DQT)[50]	NR	NR	NR	NR	NR	4 day unweighed food diary	Completed at home within the same two weeks	Total DQT score; DQT subscales; Nutrients from food diaries	Pearson's correlation	• Total DQT score with %E sfa, r= -0.50; fibre (g) r= 0.56; omega-3(mg) r= 0.33 (p<0.05). • Fibre subscale with fibre (g), r= 0.42; fat subscale with % E sfa, r= 0.49; omega-3 subscale with omega-3 (mg), r= 0.37 (p<0.05). • No correlations for TF, vitamin C or salt subscales.
Bailey Elderly Food Screener (B-Elder) [28, 51]	NR	NR	NR	NR	NR	1) 4 x 24 hour recall, Anthropometrics, Concentration biomarkers, Biomarkers of preclinical disease	4 to 6 weeks	Factor scores (dietary patterns), MAR and nutrients from recalls; biomarkers	Pearson's correlation	1) 2 patterns identified: pattern 1= "prudent dietary score"; pattern 2= "Western dietary score". Pattern 1 correlations: • With nutrients: MAR, r= 0.37; sfa (g), r= -0.25; CHO (g), r= 0.19; fibre (g) r= 0.45, p<0.001; protein (g) r= 0.25; omega-3s (g), r= 0.16; p<0.05;TF (g), r= -0.20 With biomarkers: HDL-C, r= 0.17; TGs, r= -0.15; WC r=-0.18, p<0.001 Pattern 2 correlations: •With nutrients: Sugar (g), r= 0.2; protein (g), r= -0.26; fibre (g), r= -0.20, p<0.05. With biomarkers: Serum B12 (mg), r= -0.19
	NR	NR	NR	NR	NR	2) 4 x 24 hour recall, Anthropometrics, Dietary index, Concentration biomarkers	4 to 6 weeks	3 categories of risk from screener - at risk, possible risk, not-at-risk; Nutrients, MAR and HEI-2005 from recalls		2) • At risk group reported significantly higher consumption of TF, sfa, transF and lower consumption fibre and protein. HEI and MAR were lowest in at risk group (corrections made for multiple comparisons, p<0.05) • Calculated sensitivity = 83%; specificity = 75% accuracy = 79%; positive predictive value =75%.

Table 2: Summary of reliability and relative validity

Tool	Test retest reliability					Validity				
	Retest time	**Variables**	**Test**	**Results**	**Internal reliability**	**Reference measure**	**Time between tests**	**Variables**	**Test**	**Results**
Food Behaviour Checklist - text version (FBC-T)[10, 52]	1) NR	NR	NR	NR	1) Subscale ranges, α = 0.28 (fat and cholesterol) to 0.79 (fruit and vegetables) (Spearman's) rho = 0.85 (food security)	1) 3 x 24 hour recall, Concentration biomarkers	Same visit	Subscales and individual items from screener Plasma carotenoids; HEI score, nutrients and food groups from recall	Spearman's correlation	Multiple comparisons made with subscales/individual items and nutrients, HEI score, food groups, serum carotenoids Subscales • %E TF, r= -0.25, p<0.01 ('diet quality') • HEI, from r= 0.20, p<0.05 ('fat and cholesterol', 'food security') to from r= 0.32, p<0.001 ('diet quality'). • Serum carotenoids, r= 0.28, p<0.05 ('fat and cholesterol') to r= 0.48, p<0.001 ('diet quality') Individual items • Nutrient/ food groups, from r= 0.20, p<0.05 ('one kind of fruit' with vitamin C; FBC-T servings of fruit / vegetables with HEI; 'use nutrition labels' with fibre'; 'worry about food running out' with recall servings of fruit) to r= 0.41, p<0.001 ('use nutrition labels' with vitamin A). • Serum carotenoids, from r= 0.27 p<0.05 ('fruit and vegetables as snacks') to r= 0.48 p<0.001 ('do you eat low-fat instead of high-fat foods')
	2) 3 weeks	Individual items	Spearman's correlation	From r= 0.35, p<0.05 (do you eat more than one type of fruit /day) to 0.83, p<0.001 (do you drink regular soft drinks).	2) Subscale ranges, α= 0.61 (diet quality) to 0.80 (F+V)	2) As above	As above	As above		• 17 items did not correlate and were removed.
Food Behaviour Checklist - visual version, Spanish translation (FBC-SV)[44]	3) 3 weeks	Total score individual items	ICC, Spearman's correlation	Total score, r= 0.71, p<0.001; Subscales from r= 0.48 (food security) to 0.78, p<0.001 (dairy/calcium) Individual items, from r= 0.35, p<0.01 (more than 2 servings vegetables at a main meal) to 0.79, p<0.0001 (rate eating habits). ICC total scale = 0.75 Subscales from 0.26 (sweetened beverages) to 0.80 (F+V). Individual items from 0.34 (servings of fruit and more than 2 servings vegetables) to 0.81 (drink milk).	Total score, α = 0.75 Subscales from, α = 0.49 (diet quality) to α = 0.80 (F+V) 2 item subscales r= 0.42 (dairy) and r= 0.26 (sweetened beverages) (n=154)	3) 3 x 24 hour recall	Unclear but FBC completed at the same time as second or third recall	Subscales and individual items from screener Nutrients and cups of F+V from recalls	Spearman's correlation	Multiple comparisons made with subscales / individual items and nutrients/cups of F+V Behavioural subscales • 'dairy/calcium', from r= 0.25, p<0.05 (vitamin A (RE)) to r= 0.43, p<0.001 (calcium (mg)); • 'food security' with USDA food security scale r= 0.42, p<0.001; • 'diet quality', from r= -0.23, p<0.05 (MyPyramid grains oz) to r= -0.33, p<0.01 (% E transF); • 'fast food' (higher score represents lower intake) with vitamin A and B-12, r= 0.23, p<0.05; • 'sweetened beverages' (higher score represents lower intake) from r= -0.33 (% E TF) to r=-0.41, p<0.001 (total sugar (g)). • No correlation for F+V subscale Individual items • Nutrients, from r= -0.21, p<0.05 ('red meat or pork yesterday' (higher score represents lower intake) with Dchol, mg), to r=0.43, p<0.001 ('drink milk' with vitamin D)

Table 2: Summary of reliability and relative validity

Tool	Test retest reliability					Validity				
	Retest time	**Variables**	**Test**	**Results**	**Internal reliability**	**Reference measure**	**Time between tests**	**Variables**	**Test**	**Results**
Healthy Eating Vital Signs (HEVS)[8,53]	1) NR	NR	NR	NR	NR	1) Anthropometrics	Same visit	Subscales from HEVS; BMI	Linear regression analysis	• A 0.61 increase in BMI was associated with an additional sugary drink in a typical day (0.17 to 1.04, p<0.001) but no association in 1-day recall. • A 0.91 reduction in BMI was associated with a 1-day increase in physical activity in a typical week (-1.39 to -0.44, p<0.001) with no association in 1-day recall.
	2) NR	NR	NR	NR	2) Total score, α = 0.49	Block Food Frequency Questionnaire	1 week	Individual items from HEVS Items and nutrients from FFQ	Pearson's correlation	• 'Multiple items' on HEVS with FFQ items/nutrients, from r= 0.30 (1-day and typical 'eating fast food' with TF (g); 1-day and typical F+V questions with fibre (g), p<0.05) to r= 0.5 (1-day 'eating fast food' with transF (g), p<0.001).
Latino Dietary Behaviors Questionnaire (LDBQ)[54]	NR	NR	NR	NR	Baseline total score α= 0.47; 12m α= 0.48. Healthy dietary change subscale had the strongest baseline and 12 month internal consistency (α= 0.60 and α= 0.58)	3 x 24 hour recall	Unclear but both dietary measures collected at baseline and 12 months	Total score, subscale scores and change scores from LDBQ; Nutrients, proxy 'behavioural' items and change scores from recalls (baseline and 12 months) Baseline clinical measures	Correlation	Baseline: • Total LDBQ score with sodium (mg) r= -0.25, and energy (kcal) r= -0.34, p<0.01 (no other correlations with nutrients) • Subscales, from r= 0.15, p<0.05 ('fat consumption' with %E transF) to -0.39, p<0.01 ('healthy dietary change' with energy, kcal) • Clinical measures, r= -0.16, p<0.05 (HbA1c) to r= 0.24, p<0.01 (diastolic blood pressure). 12 months: • Total LDBQ score, from r= -0.16, p<0.05(%E transF) to r= 0.37, p<0.01 (%E protein) • Subscales, r= -0.15 ('artificial sweeteners' with % E sfa) to r= -0.43, ('healthy changes' with %E TF), p<0.01.
									Independent samples t-test	• Sensitivity to change: LDBQ showed greater change over time in the intervention group (n=67; 7.10 (5.53)) compared with control group (n=75; 3.36 (5.12)), p<0.001.
PrimeScreen[55]	2 weeks	Nutrients and food group consumption derived from PrimeScreen	Spearman's correlation	For foods and food groups r= 0.50 (other vegetables) to 0.87 (adding salt) (no p values); for nutrients r= 0.59 (lutein/zeaxanthin) to 0.86 (vitamin A with supplements) (no p values). No difference across demographic strata.	NR	Willet's SFFQ	2 - 4 weeks	Nutrients and food groups calculated from both Primescreen and FFQ Plasma carotenoids and plasma vitamin levels	Pearson's correlation, Spearman's correlation	• Food groups, from r= 0.36 ('other vegetables') to 0.82 ('whole eggs'). • Nutrients, from r= 0.43 (iron) to 0.74 (vitamin E with supplements). • Plasma levels, PrimeScreen nutrients with vitamin E, r= 0.33; beta-carotene, r= 0.43; lutein/zeaxanthin, r= 0.43 (for the SFFQ these were 0.19, 0.43, 0.34 respectively). • Specificity for <3/day F+V = 67%; Sensitivity for 5/ day F+V = 73%. Sensitivity for >10% E sfa was 81% and specificity 66%. (no p values given)

Table 2: Summary of reliability and relative validity

Tool	Test retest reliability							Validity		
	Retest time	**Variables**	**Test**	**Results**	**Internal reliability**	**Reference measure**	**Time between tests**	**Variables**	**Test**	**Results**
Rapid Eating Assessment for Patients (REAP)[29]	1 week	Total score; individual items	Correlation	Total score, r= 0.86, p<0.001; For individual items from r= 0.79 (type of ice cream) to 0.33 (servings of fruit and vegetables) (p<0.001)	NR	Food diaries (unknown), Women's Health Initiative FFQ	Sample 2: 1 week (sample 1 not stated)	Study 1: REAP total score and subscales Total HEI and HEI subscales from diaries. Study 2: Modified REAP subscales; foods/nutrients from FFQ	Correlation	Study 1 • Total scores, r= 0.49, p= 0.007 • Subscales: from r= 0.31, p= 0.04 (variety subscales) to r= 0.55, p<0.001 (fat subscales) Study 2 • REAP energy subscale with energy (kcal), r= -0.44, p<0.001 • Other subscales, from r= 0.30, p=0.024 (fruit servings) to r= -0.62, p<0.001 (alcohol)
Rapid Eating Assessment for Patients short form (REAP-S)[14]	NR	NR	NR	NR	NR	Block 1998 semi quantitative food frequency questionnaire	Unclear	REAP Food groups; Food groups, TF, fibre (g),Dchol (mg) sugar (g) from FFQ	Pearson's correlation	• Food groups, from r= -0.38 (added fat servings) to r= 0.51 (fruit servings), p<0.001. • REAP-S food servings with FFQ nutrients, from r= -0.20, p= 0.034 ('fish, poultry and meat servings' with Dchol (mg)) to 0.52, p<0.001 ('vegetable servings' with fibre (g)).
Short Diet Quality Screener (sDQS)[13] Developed in the same population as the Brief Mediterranean Diet Screener (bMDSC)	NR	NR	NR	NR	NR	10 x 24 hour recall	24 hr recalls >1 year. (bMDSC first, then 1 week later sDQS)	Total score from sDQS (DQI); DQI from recalls	Pearson's correlation, ICC, LOA Mann Whitney U Gross misclassification	• DQI, r= 0.61 (no p value) • sDQS DQI mean vs 24hr DQI mean = 39.3 (2.8) vs 35.5 (2.8), difference 3.82 (95% CI, 3.33, 4.31). LOA = 96% to 126%. ICC = 0.32. • 48.5% participants classified in the same tertile, 3.9% in the opposite tertile.
Adherence to the Mediterranean diet										
Brief Mediterranean Diet Screener (bMDSC)[13] Developed in the same population as the Short Diet Quality Screener (sDQS)	NR	NR	NR	NR	NR	10 x 24 hour recall	24 hr recalls >1 year. bMDSC first, then 1 week later sDQS	Total and subscales (ANTOX-S and mMDS) from bMDSC; ANTOX-S and mMDS scores derived from recalls	Pearson's correlation, ICC LOA Mann Whitney U Gross misclassification	• bMDSC mMDS with 24hr mMDS, r= 0.40 • bMDSC ANTOX-S with 24hr ANTOX-S, r= 0.45 (no p values) • bMDSC mMDS mean = 18.3 vs 24hr mMDS mean = 20.7, difference= - 2.44 (95% CI -3.01, -1.82). Mean differences for the ANTOX-S was zero. • For the mMDS 44% participants classified in the same tertile, 11% in the opposite tertile. • LOA = 61% to 118%, ICC = 0.30. • For the ANTOX-S 50% in the same tertile with 9% in the opposite tertile. LOA = 59% to 144%, ICC= 0.45.

Table 2: Summary of reliability and relative validity

Tool	Test retest reliability					Validity				
	Retest time	**Variables**	**Test**	**Results**	**Internal reliability**	**Reference measure**	**Time between tests**	**Variables**	**Test**	**Results**
Mediterranean Diet Adherence Score (MEDAS)[56]	NR	NR	NR	NR	NR	Anthropometrics, Biomarkers of preclinical disease, PREDIMED FFQ	Unclear - baseline	PREDIMED score and individual items from MEDAS PREDIMED score from FFQ, individual items and nutrient intakes from FFQ; anthropometrics biomarkers	Pearson's correlation General linear modelling ICC, LOA Kappa statistics	• Total PREDIMED scores, r= 0.52, p<0.001. • Absolute ICC = 0.52 for men, 0.51 for women, p<0.001. • Associations found for CVD risk factors and MEDAS for BMI (β= -0.146, p<0.001) and waist circumference (β= -0.562, p<0.001) with smaller associations for lipids and fasting blood glucose. Individual items • Associations between nutrients / foods on the FFQ and PREDIMED quintiles as derived by MEDAS were in the expected direction, except for vitamin E where there was no association. For example the 1st quintile consumed 155g fruit vs 180g for the 5th quintile (p<0.001). • Kappa scores for individual items ranged from 0.03 ('consuming sauces with tomatoes') to 0.81 ('wine'), with a mean of 0.43 (moderate): 21.4% of items showed poor agreement between screener and the FFQ with 21.4% of items good or excellent. • 47.9% of individuals were grouped into the same PREDIMED tertile on MEDAS and FFQ; 8.6% grouped in opposite tertiles.
Total Fat										
Dutch fat consumption questionnaire (D-Fat1)[57]	1 year	Total score; tertiles	Pearson's correlation Gross misclassification	r= 0.71 (no p value); 3.9% of participants were classified in opposite fat consumption tertiles.	NR	7 day un-weighed food diary	1 week - 1 month	Total Dutch fat score TF (g) from diaries	Pearson's correlation Unweighted kappa statistics Gross misclassification	• Dutch Fat score with TF (g), r= 0.59 (no p value) • Kappa = 0.42 with 2 categories and 0.25 with 3. • Gross misclassification = 15.4%
Fat Related Diet Habits Questionnaire / Kristal's Food Habits Questionnaire – 20 items (FRDHQ)[11]	1) 3 months	Total FRDHQ score; Behavioural subscales	Correlation	Total FRDHQ score, r= 0.87; for subscales from r= 0.67 (replace high fat foods with naturally low fat foods) to 0.90 (avoid fat as a seasoning), (no p values).	1) Total FRDHQ score, α= 0.62; subscales range α= 0.54 (replace high fat foods with naturally low fat foods) to 0.76 (avoid meat)	1) 8 day food diaries (unknown), Modified Block/National Cancer Institute FFQ	After the 1st food diary was returned	Total FRDHQ score and behavioural subscales from KFHQ; %E TF from diet records and FFQ;	Adjusted correlation	• Total FRDHQ score with %E TF, r= -0.60 p<0.001 Subscales • %E TF, from r= -0.29, p<0.01 ('avoid meat') to r= 0.50, p<0.001, ('avoid fat as seasoning'). • Linear relationship with %E TF and 'avoid fat as seasoning', 'substitution of high fat foods with manufactured low fat alternatives' and 'replace high fat foods with naturally low fat foods'. • On 'avoiding meat' those scoring <2.0 (n=55) had a higher %E TF than those with scores >2 (n=40); on the 'modify high fat food (trimming fat/skin from meat)' those scoring 4 (n=59) had a lower %E TF than other groups (30.6% vs approx 36%). (No statistical tests or p values are described). • In a multiple regression model predicting %E TF from all components summary R squared = 0.47.

Table 2: Summary of reliability and relative validity

Tool	Test retest reliability					Validity				
	Retest time	**Variables**	**Test**	**Results**	**Internal reliability**	**Reference measure**	**Time between tests**	**Variables**	**Test**	**Results**
Fat Related Diet Habits Questionnaire / Kristal's Food Habits Questionnaire – 20 items (FRDHQ)[46, 21]	2) NR	NR	NR	NR	2) Total FRDHQ score, α= 0.73; subscales range, α = 0.13 (replace high fat foods) to 0.53 (make modifications to meat prep). When used as a behavioural checklist α= 0.70. (Item-scale correlations also tested)	2) Ontario Health Survey Food Frequency Questionnaire		Total FRDHQ score and subscales; % E TF and total energy (kcal) from FFQ	Adjusted correlation	• Total FRDHQ score with %E TF, r= -0.24, p<0.001. Subscales • %E TF, r= -0.12, p<0.05 ('substitute low-fat for high fat') to -0.24, p<0.001 ('avoid fat as a seasoning'). • No significant correlations for energy with total score or subscales. • Confirmatory factor analysis showed discrepancy between the hypothesised structure and the actual responses (likelihood ratio (160) = 256.98, p= 0.001. Some items were not related to hypothesised factor e.g loadings of less than 0.3 for replace high fat food subscale. • When the tool was used as a behavioural checklist total score with %E TF, r= -0.27 (p<0.001).
	3) i) NR	NR	NR	NR	NR	3) i) Block / National Cancer Institute FFQ (B-FFQ) Block fat screener (B-FS)	2 years (B-FHQ) Same visit (B-FHS)	Total score from FRDHQ; %E TF from B-FFQ; Total score from B-FS BMI; TChol	Correlation	• Total FRDHQ score with %E TF r=0.48, p<0.01 • Total FRDHQ score with B-FS score, r= 0.61, p<0.01 • Total FRDHQ score with BMI, r= 0.1, p<0.01 • No correlation with TChol
	3) ii) 3 months	Total score	Correlation	All participants r=0.59, p<0.01; UC only r=0.56, p<0.01	NR	3) ii) 4 day food diaries	After return of food diaries	Total score from FRDHQ; Energy (kcal), % E TF, kcal from diary; BMI, HbA1c (%), TChol (mmol/l) at baseline	Correlation	• Total FRDHQ score with energy (kcal), r=0.27, p<0.01 • Total FRDHQ score with %E TF, r= 0.44, p<0.01 • Total FRDHQ score with TChol, r= 0.19, p<0.05 • Total FRDHQ score with HbA1c, r= 0.32, p<0.01 • Total FRDHQ score with BMI, r= 0.22, p<0.05
								Follow up scores, adjusted for baseline used to calculate responsiveness to change	'Responsiveness' = mean difference between UC (n=39) and intervention (n=50)	Baseline: • UC mean FRDHQ score = 2.14 • Intervention mean FRDHQ score = 2.16 3 months • UC mean FRDHQ score = 2.16 • Intervention mean FRDHQ score = 1.97 • FRDHQ responsiveness = 0.4

Table 2: Summary of reliability and relative validity

Tool	Test retest reliability							Validity		
	Retest time	**Variables**	**Test**	**Results**	**Internal reliability**	**Reference measure**	**Time between tests**	**Variables**	**Test**	**Results**
Fat Related Diet Habits Questionnaire / Kristal's Food Habits Questionnaire – 24 items (FRDHQ)[47]	4) 9 months	Total FRDHQ score; Behavioural subscales	Pearson's correlation	Total FRDHQ score, r= 0.74 p<0.01 Subscales from, r= 0.48 (replace with fruit) to 0.68 (avoid fat), (p<0.01)	Total FRDHQ score, α = 0.83 Subscales from, α= 0.47 (replace fat) to 0.76 (substitute fat)	4 day un-weighed food diary	baseline to 9 months	Total FRDHQ score and subscales from KFHQ; Energy (kcal), %E TF and TF (g) from diaries(pre and post intervention)	Correlation	Pre-intervention • Total FRDHQ score with energy (kcal), r= 0.43, p<0.05; TF (g), r= 0.52 with % E TF, r= 0.47, p<0.01. Subscales • r= 0.35 ('avoid fat' and 'substitute fat' with energy) to r= 0.43 ('modify meats' with TF (g)), p<0.05. No subscale correlated with %E TF. • 'Replace with fruit' subscale did not correlate with the nutrient estimates. Post-intervention: • Total FRDHQ score with TF (g) r= 0.46, p<0.01. No other correlation for total FRDHQ score. Subscales • r= 0.21 ('modify meats' with energy), p<0.05 to 0.47 ('substitute fat' and TF (g)), p<0.01.
Short Fat Questionnaire (SFQ)[56]	7 - 9 months	Total score	Pearson's correlation	r= 0.85 (95% CI, 0.69-0.93)	NR	179-item CSIRO FFQ	Unstated, possibly same visit	Total score from SFQ %E TF, % E sfa; pufa:sfa from FFQ	Correlation Misclassification	• Total score with %E TF, r= 0.55 (CI 0.39 to 0.68), %E sfa, r= 0.67 (CI 0.54 to 0.77) and pufa:sfa, r= -0.44 (CI -0.60 to -0.26). • 38% of participants were in the same quartile for %E TF, 46% differed by one quartile • 43% same quartile for %E sfa, 44% differed by one quartile.
Sister Talk Food Habits (short form)[3]	NR	NR	NR	NR	Total score, α = 0.79	Anthropometrics, 91 item SisterTalk FFQ	Same time	Change in BMI Change in short Sister Talk	Pearson's correlation Bootstrapping,	Post intervention • Change in SisterTalk with change in BMI, r= 0.17 (95% CI 0.02, 0.39) During maintenance • Change in SisterTalk with change in BMI, r= 0.28 (95% CI 0.02, 0.50), p<0.001. • Total change in SisterTalk with total change in BMI, r= 0.35 (95% CI 0.08, 0.58), p<0.05 • (Correlations were not significantly different between short and long Sister Talk FHQ.)
Starting the Conversation (STC)[39]	4 months	Total score; individual items	Pearson's correlation	Total score r= 0.66, p<0.05 Individual items from, r= 0.4 to r= 0.62 (no details), p<0.05.	(Pearson's) Individual items correlated with the summary (r= 0.39 to 0.59, p<0.05).	NCI Percentage energy from fat (Pfat) screener	Both tools completed at the same visit	STC total score and change in STC total score Pfat score and reduction in TF from Pfat	Pearson's correlation	• Baseline STC total score and Pfat score, r= 0.39, p<0.05; • Change in STC score and reduction in Pfat TF, r= 0.22, p<0.05.

Table 2: Summary of reliability and relative validity

Tool	Test retest reliability							Validity		
	Retest time	**Variables**	**Test**	**Results**	**Internal reliability**	**Reference measure**	**Time between tests**	**Variables**	**Test**	**Results**
Specific dietary fats and / or dietary cholesterol										
Dietary Fat Quality Assessment (DFA)[60]	2 - 4 weeks	Dietary fats quantified from DFQA servings	ICC	ICC range = 0.48 to 0.59 for dietary fats (no CIs given)	NR	Fred Hutchinson Cancer Research Center FFQ	2 - 4 weeks	Total DFQA; Quantified fat and cholesterol intakes from DFQA and FFQ pufa:sfa from FFQ	Spearman's correlation Gross misclassification	• Total DFQA score with PUFA:SFA ratio, r= 0.4, p<0.001. • DFQA with FFQ fat estimates, r= 0.54 (Dchol, (g)) to r= 0.66 (sfa, (g)), p<0.001. • DFQA classified 39% (mufa, (g)) to 55% (sfa, (g)) of participants into the same nutrient quartile as the FFQ and 80% to 87% into adjacent quartiles. 2% of participants were grossly misclassified for sfa, mufa, omega-3s and dchol.
Heart Disease Prevention Project Screener (HDPPS)[61]	3 - 4 months	Total score	Mean scores, gross misclassification	"Only small differences between occasions (20 of the men had a difference of 2 points or less) and the mean scores for the group were identical on each occasion" (p365)	NR	3 day food diaries (unknown)	Around the same time	Total screener score; Mean sfa (g) from food diary	Correlation, Independent samples t-test	• Total score with sfa (g) r=-0.30, p<0.05 • Estimated mean sfa (g) for 34 men with scores of <15 was 53.4g vs 41.2g for those with scores of >16 (n=34), p<0.001
MEDFICTS[12, 62]	1) NR	NR	NR	NR	NR	1) 3 x food diaries (unknown)	"recent"	Total score; %E TF, %E sfa and Dchol from diaries	Pearson's correlation	Sample 1 (n=22) • Total score with %E TF, r= 0.79 (p<0.002), %E sfa, r= 0.60 (p<0.003), Dchol, r= 0.71 (p= 0.001) Sample 2 (n=26) • Total score with %E TF, r= 0.54 (p= 0.009), Dchol, r= 0.39 (p= 0.051).
								Total score; consumption of Step 1, 2 and 3 diets from food diaries as assessed by a dietitian		• Pre-existing food diaries: "Medficts scores correctly identified the 11 patients consuming a Step 1 diet…the 2 patients consuming a Step 2 diet and the 3 patients consuming an average American diet" p85
	2) NR	NR	NR	NR	NR	2) Reduced Block FFQ		Total score and diet classification (high fat, step 1, step 2) from Medficts Energy (kcal), %E TF,%E sfa and Dchol from FFQ	Spearman's correlation Kappa statistics ROC curve analysis	• Total score, r= 0.52, p<0.0001 (%E TF and %E sfa), Dchol, r= 0.55, p<0.0001. • Identified as high fat diet: FFQ identified 76.2% vs MEDFICTS identifying 17.7% of this group. • Recommended Medficts cut offs correctly identified 23.3% high fat diets and 19.2% Step 1 diets. No agreement for diet steps between FFQ and Medficts, kappa = 0.036. • ROC curve analysis showed that a single cut off of 38 gave sensitivity of 75% and specificity of 72% and modest agreement with FFQ, kappa = 0.39, p<0.001 for a high fat diet.

Table 2: Summary of reliability and relative validity

Tool	Test retest reliability					Validity				
	Retest time	**Variables**	**Test**	**Results**	**Internal reliability**	**Reference measure**	**Time between tests**	**Variables**	**Test**	**Results**
MEDFICTS[63, 64]	3) NR	NR	NR	NR	NR	3) Arizona FFQ	Mean of 52 days	Total score from Medficts %E TF; %E sfa; Dchol from FFQ	Spearman's correlation Chi squared Independent samples t-test ROC curve analysis	• Total score with %E TF, r= 0.30, p<0.001 • Identified high fat diet: FFQ identified 71.2% vs Medficts identified 50.5%. 59.8% identified on both tools. • Dichotomized Medficts score >30% energy from fat, chi squared = 8.19, p<0.01; sensitivity of Medficts for >30% energy from fat = 57.3%; specificity = 66%. • Positive predictivity (ie classifying high fat diets as the FFQ) = 80.6%, negative predictivity (classifying low fat diets as the FFQ) = 38.5%.ROC curve analysis indicated Medficts was better than chance (p= 0.03).
	4) NR	NR	NR	NR	NR	4) Block 98 FFQ	Same visit	Total score and classified as adherent/ non adherent to TLC diet from Medficts %E TF, %E sfa and Dchol (mg) from FFQ	Pearson's correlation Chi squared Kappa statistics	• Total score with %E sfa, r= 0.52, %E TF, r= 0.31, Dchol r= 0.54, p<0.0001. • Medficts categorised 44.9% of participants as adherent to the TLC diet. FFQ categorised 4.2% as adherent. • Categorical agreement • Overall, k= 0.08, p<0.001; <7% sfa; k= 0.13, p<0.001; <30% TF, k= 0.16, p<0.001; <200mg Dchol, k= 0.34, p<0.001. • Sensitivity, adherent to TLC diet= 85.7% of the time, and specificity, non-adherent = 56.9% of the time. • Specificity lower for women (48.4%) vs men (72.9%) p<0.001. Optimal cut off point < 25 improved specificity to 82.5% and sensitivity of 76.2% overall but men and women were different with men having an optimal cut off < 37 (specificity of 80% and sensitivity of 78.3%) and women an optimal cut off < 20 (specificity < 83.8%, sensitivity < 75%). No difference seen for ethnicity for sensitivity or specificity.
NLSChol Questionnaire[58]	30 days	3 groups from NLSChol 3 groups derived from diet history.	ICC, Paired Wilcoxon rank score Percentage agreement in classification	ICC = 0.89 (no Cis); Agreement in classification = 85% (17 patients), 15% (3 patients) moved up or down a group. Comparison of medians was not significant, p= 0.52.	Total score, α = 0.69	Diet history		3 groups from NLSChol 3 groups derived from diet history.	Pearson's correlation Kappa statistics Bowker's test of symmetry	• Group classification, r= 0.3, p= 0.029 • Agreement of 72% between dietitian classification and NLSChol score, kappa = 0.48 (0.10; 0.69). • Bowker's test of symmetry was not significant.

Table 2: Summary of reliability and relative validity

Tool	Test retest reliability							Validity		
	Retest time	**Variables**	**Test**	**Results**	**Internal reliability**	**Reference measure**	**Time between tests**	**Variables**	**Test**	**Results**
Northwest Lipid Research Clinic Fat Intake Score (NWFIS)[20]	2-3 weeks; 6-8 weeks	Total score	Pearson's correlation	Retest after 2-3 weeks r= 0.88 for men, 0.90 for women (p<0.001) After 6 - 8 weeks r= 0.76 for men and 0.78 for women (p<0.001)	NR	4 day un-weighed food diary, Biomarkers of preclinical disease	NR	Total NWFIS and change in FIS TF, sfa, Dchol (adjusted and not adjusted for energy), Keys score, RISCC score, change in nutrients, change in Keys score and change in RISCC score from diaries Change in plasma cholesterol	Pearson's correlation	Baseline • Total NWFIS with %E TF, r= 0.49, %E sfa, r= 0.44, Dchol mg/1000kcal, r= 0.46, Keys score, r= 0.46, RISCC, r= 0.53 p<0.001. 18 months • Total NWFIS with %E TF, r= 0.55, %E sfa, r= 0.64, p<0.001 and Dchol mg/1000kcal, r= 0.30, p<0.01, Keys score, r= 0.58 and RISCC, r= 0.56, p<0.001 • Change in NWFIS with change in TF (g), r= 0.38 (men) and r= 0.40 (women), in sfa (g), r= 0.42 (both), in Dchol (mg), r= 0.32 (men) and 0.52 (women), in Keys score, r= 0.38 (men) and 0.48 (women), in RISCC, r= 0.39 (men) and 0.51 (women) p<0.001.
Rate Your Plate (RYP)[15]	NR	NR	NR	NR	NR	Willett SFFQ	Same visit	Total score from RYP; dietary fats and Dchol from FFQ	Pearson's correlation	• Total score, r= -0.28 (% E TF, less trimmed fat) to r= -0.48 (% E sfa), p<0.05.
Total and saturated fat and free sugar										
Dietary Fat and Free Sugar Short Questionnaire (DFFQA)[69]	158 (10) days	Total score	ICC	ICC = 0.83 (95% CI 0.66 - 0.91)	Total score, α = 0.76.	Anthropometrics, Commonwealth Scientific and Industrial Research Organisation Food Frequency Questionnaire 4 day un-weighed food diary	Same visit for FFQ, shortly after for food diary	Total DFFS score; DFFS subscales; Nutrients from FFQ	Spearman's correlation Independent samples t-test	• DFFS with food diary nutrients, from r= 0.35 (energy), p<0.05 to r= 0.46 (% E sfa), p<0.01. • DFFS with FFQ nutrients, from r= 0.40 (energy) to r= 0.71 (% E sfa), p<0.01. • Subscales with nutrients, from r= 0.33 (fat subscale with diary %E sfa), p<0.05 to r= 0.68 (fat-sugar subscale with FFQ %E sfat). • For DFS scores < 60 mean %E TF= 28.56 vs DFS score > 60 mean %E TF= 33.51 (3.87), p < 0.01; DFS scores < 60 mean %E free sugars= 7.41 (4.54) vs DFS > 60 mean %E free sugars= 11.39 (6.15), p < 0.05.
Dietary fats and fibre										
Dietary Instrument for Nutrition Education (DINE)[5]	NR	NR	NR	NR	NR	4 day un-weighed food diary	5 days	DINE fat score; DINE fibre score; Fat and fibre intake from diaries	Pearson's correlation Weighted kappa Percentage agreement of classification Gross misclassification	• DINE fat score, from r= 0.28 (%E TF) to 0.57 (sfa, (g)); fibre score with fibre (g) r= 0.46; DINE unsaturated fat score with pufa:sfa ratio, r= 0.43, p<0.001. • Weighted kappa = 0.38 for TF (g) and 0.30 for fibre (g). • Exact agreement of categorization was 53% for TF (g), 52% for fibre (g). Gross misclassification= 6% for TF (g) and 5% for fibre (g). • By tertiles 53% agreed and 7% misclassified for TF (g) and 49% agreed with 10% misclassification for fibre (g).

Table 2: Summary of reliability and relative validity

Tool	Test retest reliability					Validity				
	Retest time	**Variables**	**Test**	**Results**	**Internal reliability**	**Reference measure**	**Time between tests**	**Variables**	**Test**	**Results**
Fat and Fibre Barometer (FFB)[9]	1) 7 - 9 weeks	Total score	Pearson's correlation	r= 0.92 (95% CI, 0.89 to 0.94). No difference between men and women	1) NR	1) Geelong, meal based FFQ developed by the Deakin Institute	7-10 weeks	FFB score; TF (g), total fibre (g), % E TF, fibre (g/10MJ) from FFQ	Pearson's correlation Weighted kappa Gross misclassification	• Men: FFB score with %E TF, r= -0.33 (-0.05, -0.56); fibre (g/10MJ) r= 0.83 (0.71, 0.90). • Women: FFB score with %E TF, r= -0.75 (-0.60, -0.85), fibre (g/10MJ) r= 0.58 (0.36, 0.74). Weighted kappa • Men: %E TF = 0.39 (0.18, 0.61), fibre (g/10MJ)= 0.59 (0.42, 0.76) • Women: %E TF = 0.58 (0.41, 0.75), fibre (g/10MJ) = 0.27 (0.06, 0.48). Gross misclassification • Men: 15% for TF (g), 9% for %E TF, 6% for fibre (g), 0 for fibre g/10MJ • Women: 12% for TF (g), 2% for %E TF and 10% for both fibre variables
Fat and Fibre Diet Behaviour Questionnaire (FFDBQ)[18]	3 months 1 year	FFDB Fat score; FFDB fibre score; subscale scores	Spearman's correlation	3 months: FFDB fat score, r= 0.79; FFDB fibre score, r= 0.74, p<0.001 Subscales from r= 0.60 (modify meals to be low in fat) to 0.74 (substitute for low-fat foods), p<0.001. 1 year: FFDB fat score, r= 0.74; FFDB fibre score, r= 0.70 for fibre score, p<0.001 Subscales r= 0.53 (modify meals to be low in fat) to 0.66 (substitute high-fibre for low-fibre foods; Substitute especially manufactured low fat foods), P<0.001.	NR	FFQ was developed for the study and based on a pre-existing FFQ evaluated against diet records	Same time	FFDB Fat score; FFDB fibre score, subscale scores; %E TF, fibre/1000kcal; servings of F+V from FFQ	Spearman's correlation	• FFDB fat score with % E TF, r= 0.53, p<0.001 • FFDB fibre score with fibre/1000kcal, r= 0.50, p<0.001; with servings of F+V, r= 0.50, p<0.001. Subscales • %E TF, from r= 0.2 (replace high-fat meats) to r= 0.43 (avoid fat as a flavouring), p<0.001. • Fibre/1000kcal, from r= 0.24 (substitute high fibre for low fibre) to r= 0.43 (F+V), p<0.001.
Norweigian SmartDiet Questionnaire (N-Smart)[7]	Same day	Total score; individual items	Pearson's correlation, Percentage agreement of classification, Weighted kappa	r= 0.95 (no p value); mean agreement rate = 0.93 (range 0.85 for vegetables to 0.98 for milk); Weighted Kappa ranged from 0.75 (95% CI, 0.63 - 0.86) (vegetables) to 0.97 (95% CI, 0.94 - 1.00) (cheese).	NR	7 day weighed food diaries	8 days (screener first)	Total score, foods and food groups from screener Total calculated score, food, food groups and nutrients from diary	Pearson's correlation LOA Weighted kappa	• Total scores, r= 0.73 • Correlations with nutrients was highest for sfa (g), r=-0.59 • Kappa from 0.71 (0.56 - 0.86) for milk and 0.73 (0.60 - 0.86) for spreads to 0.42 (0.28 - 0.55) for fruit and vegetables. • Agreement ranged from 0.98 for milk to 0.38 for fish, mean agreement = 0.73 • Distribution of difference between the food diaries and tool total score, mean = 1.9 (95% CI 1.4 to 2.5), LOA = -3.8 to 7.7

Table 2: Summary of reliability and relative validity

Tool	Test retest reliability					Validity				
	Retest time	**Variables**	**Test**	**Results**	**Internal reliability**	**Reference measure**	**Time between tests**	**Variables**	**Test**	**Results**
Total fat and fruit and vegetables										
Block Fat, Fruit and Vegetable Screeners (B-F&FV)[6]	NR	NR	NR	NR	NR	100 item Block FFQ	Posted together	Meats/Snacks score and F+V scores from screener; fat nutrients, F+V servings, fibre (g) and micronutrients from FFQ;	Spearman's correlation	• Meat/snack score, from r= 0.60 (Dchol, mg) to r= 0.72 (sfa, g), p<0.0001 • F+V score (without pulses), from r= 0.41 (magnesium, mg) to r= 0.71 (fruit and vegetable servings), p<0.0001. • F+V screener (including pulses), from r= 0.46 (magnesium, mg) to r= 0.62 (fibre, g), p= 0.0001. • 89% of people low on F+V score were very low or quite low on FFQ. 12% of people scored as needing advice on fat on the screener did not need advice according to FFQ.
Hispanic Fat and Fruit and Vegetable screeners (H-F&FV)[42]	1 month	F+V score; fat score	Pearson's correlation, percentage agreement	r= 0.64 for F+V score, 0.85 for fat score, p<0.001. 84% agreement for vitamin supplement use, p<0.001	NR	NR	NR	NR	NR	NR
Fruit and/or vegetables										
Canadian Fruit and Vegetable Questionnaire (CFV-Q)[66]	NR	NR	NR	NR	NR	FFQ developed and tested by Goulet et al at Laval University	Same visit	Servings F+V from FV-Q and FFQ	Pearson's correlation ICC ROC curve analysis	• Servings, r= 0.66, p<0.001 (obese participants); r= 0.65, p<0.001 (non-obese). • ICC for obese = 0.44, for non-obese = 0.46. • Sensitivity in obese group = 88.5%, specificity = 63.6%. Positive and negative predictive values = 45% and 94%. • Non obese sensitivity = 80% and specificity = 66%, positive predictive values = 40% and negative = 92% (no difference between obese and non-obese). • ROC curve indicated that the more accurate cut-off point ≥5 servings/day vs <5 servings/day (c = 0.74).

Table 2: Summary of reliability and relative validity

Tool	Test retest reliability					Validity				
	Retest time	**Variables**	**Test**	**Results**	**Internal reliability**	**Reference measure**	**Time between tests**	**Variables**	**Test**	**Results**
Dutch fruit and vegetable questionnaire (D-F&V)[19]	1 month; 1 year	F+V intake; tertiles of intake	Spearman's correlation Percentage agreement of classification	Retest at 1 month: Total F+V, r= 0.80. Individual variables, r= 0.49 (other fruits) to 0.82 (F+V juices). Agreement classification: vegetables = 59%; 41% up or down a class fruits = 74% ; 24% up or down a class; 3% up two classes. Retest at 1 year: total F+V, r= 0.79. Individual variables r= 0.31 (other fruits) to 0.81 (total vegetables). Agreement classification, vegetables = 70%; 30% up or down a class; fruits = 57%; 40% up or down a class; 2% down two classes.	NR	Change in concentration biomarkers	Same time	F+V intake from screener; mean changes in screener score; Mean change in plasma carotenoids; Mean change in plasma vitamin C	Spearman's correlation	Baseline • Total F+V intake, from r= 0.23 (plasma B-carotene and lutein) to r= 0.37 (plasma vitamin C), p<0.01. After intervention • Changes in total F+V consumption, from r= 0.26 (change in B-carotene) to r= 0.39 (change in B-crytoxanthin), p<0.01.
Five a day screener (5-F&V) Five a day screener (5-F&V)[17,67]	1) NR	NR	NR	NR	NR	2 x 24 hour recall	1 year	Mean and median servings F+V/ day (adjusted for within person variation) from screener and recalls.	Linear regression analysis; Maximum likelihood estimates	Men • β= 0.52; median (IQR) F+V servings from the recall vs the screener = 6.6 (3.6) vs 3.7 (2.6), p<0.001 Women • β= 0.50; median (IQR) F+V servings from recall vs screener = 5.5 vs 4.2 (no IQRs quoted), p<0.001. Percentages eating 5 a day • Men = 73% for recall vs 24% for the screener • Women = 59% for recall vs 36% for the screener
	2) NR	NR	NR	NR	NR	3 x 24 hour recall	NR	Servings F+V, including and not including salad, potatoes, fried vegetables, and fruit juices from screener and recalls.	Pearson's correlation Paired sample t-test	• Total F+V servings, r= 0.50, p<0.001 • Individual items, from r= 0.27, p<0.01 ('cooked, excluding fried vegetables') to r= 0.59, p<0.001 ('fruit'). • Total F+V servings, recall vs screener = 4.1 (2.1) vs 3.3 (1.9), p<0.001. • Fruit alone recall vs screener= 0.97 vs 0.84; juice alone recall vs screener= 0.37 vs 0.56; total vegetable recall vs screener= 2.53 vs 1.80, p<0.001. No difference for fruit + juice.

Table 2: Summary of reliability and relative validity

Tool	Test retest reliability							Validity		
	Retest time	**Variables**	**Test**	**Results**	**Internal reliability**	**Reference measure**	**Time between tests**	**Variables**	**Test**	**Results**
Mainvil Fruit Habits Questionnaire (M-FRHQ)[68]	NR	NR	NR	NR	NR	Diet history	Same day (diet history last)	Daily servings of fruit from M-FRHQ and diet history	Pearson's correlation LOA, Kappa statistics Percentage of agreement Gross misclassification	• Daily servings, r= 0.57, p<0.001 (men, r= 0.67, p<0.001;women, r= 0.49, p<0.001 • LOA= -2.98 to 4.31, mean difference=0.66 (CI: 0.32,1.02) • kappa= 0.41, p < 0.001 • Percentage of agreement= 60%; 12.5% were grossly misclassified • Correctly classified 70% as achieving or not achieving ≥ 2 servings/day. • Overestimated group mean by 32%, p<0.001 • Positive predictive value was 87% (64% sensitivity), negative predictive value was 66% (88% specificity)
Short Dutch questionnaire to measure fruit and vegetables[16]	NR	NR	NR	NR	NR	7 day un-weighed food diary		Mean daily intakes of fruit and vegetables from both questionnaires	Spearman's correlation, Percentage agreement of classification	• Total F+V, r= 0.43; total fruit, r= 0.51; total vegetables, r= 0.35. • 36.8% misclassified for total fruit intake; 22.5% for total vegetable intake. 7 people (14%) were classified as meeting recommendations for total F+V intake on screener who were not on FFQ.

%E = % energy; TF= Total fat; sfa=saturated fat; pufa=poloyunsaturated fat; mufa=monounsaturated fat; transF=trans fat; Dchol=dietary cholesterol; F+V=Fruit and vegetables; MAR=Mean Nutritional Adequacy Ratio; HEI=Healthy Eating Index; ANTOX-S= Antioxidant Score; mMDS= modifided Mediterranean Diet Score; DQI= Diet Quality Index; RISCC=Ratio of ingested saturated fat and cholesterol to calories; TLC=Adult Treatment Panel III Therapeutic Lifestyle Changes diet
BMI=body mass index WC=waist circumference (cm); TChol=total cholesterol; HDL-C= HDL-cholesterol (mmol/l); TG=triglycerides (mmol/l), DBP=diastolic blood pressure
ICC=Intra-class correlation; LOA=Limits of agreement; PCA=Principle Component Analysis; OR=Odds Ratio; CI=Confidence interval
FFQ=Food frequency questionnaire