British Journal of Nutrition (2016), **116**, 1787–1798
doi:10.1017/S0007114516003755

Development and validation of empirical indices to assess the insulinaemic potential of diet and lifestyle

Fred K. Tabung[1,2]*†, Weike Wang[1,2]†, Teresa T. Fung[1,3], Frank B. Hu[1,2,4,5], Stephanie A. Smith-Warner[1,2], Jorge E. Chavarro[1,2,4,5], Charles S. Fuchs[4,5,6], Walter C. Willett[1,2,4,5] and Edward L. Giovannucci[1,2,4,5]

[1]*Department of Nutrition, Harvard T.H. Chan School of Public Health, Boston, MA 02115, USA*
[2]*Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA 02115, USA*
[3]*Department of Nutrition, Simmons College, Boston, MA 02115, USA*
[4]*Channing Division of Network Medicine, Brigham and Women's Hospital, Boston, MA 02115, USA*
[5]*Department of Medicine, Harvard Medical School, Boston, MA 02115, USA*
[6]*Department of Medical Oncology, Dana–Farber Cancer Institute, Boston, MA 02115, USA*

(Submitted 9 June 2016 – Final revision received 16 September 2016 – Accepted 29 September 2016 – First published online 8 November 2016)

Abstract

The glycaemic and insulin indices assess postprandial glycaemic and insulin response to foods, respectively, which may not reflect the long-term effects of diet on insulin response. We developed and evaluated the validity of four empirical indices to assess the insulinaemic potential of usual diets and lifestyles, using dietary, lifestyle and biomarker data from the Nurses' Health Study (NHS, *n* 5812 for hyperinsulinaemia, *n* 3929 for insulin resistance). The four indices were as follows: the empirical dietary index for hyperinsulinaemia (EDIH) and the empirical lifestyle index for hyperinsulinaemia (ELIH); the empirical dietary index for insulin resistance (EDIR) and the empirical lifestyle index for insulin resistance (ELIR). We entered thirty-nine FFQ-derived food groups in stepwise linear regression models, and defined indices as patterns most predictive of fasting plasma C-peptide, for the hyperinsulinaemia pathway (EDIH and ELIH), and of theTAG:HDL-cholesterol ratio, for the insulin-resistance pathway (EDIR and ELIR). We evaluated the validity of indices in two independent samples from NHS-II and Health Professionals Follow-up Study (HPFS) using multivariable-adjusted linear regression analyses to calculate relative concentrations of biomarkers. The EDIH is comprised of eighteen food groups; thirteen were positively associated with C-peptide and five were inversely associated. The EDIR is comprised of eighteen food groups; ten were positively associated with TAG:HDL-cholesterol and eight were inversely associated. Lifestyle indices had fewer dietary components, and included BMI and physical activity as components. In the validation samples, all indices significantly predicted biomarker concentrations – for example, the relative concentrations of the corresponding biomarkers comparing extreme index quintiles in the HPFS were EDIH, 1·29 (95 % CI 1·22, 1·37); ELIH, 1·78 (95 % CI 1·68, 1·88); EDIR, 1·44 (95 % CI 1·34, 1·55); and ELIR, 2·03 (95 % CI 1·89, 2·19); all $P_{trend} < 0·0001$. The robust associations of these novel hypothesis-driven indices with insulin response biomarker concentrations suggest their usefulness in assessing the ability of whole diets and lifestyles to stimulate and/or sustain insulin secretion.

Key words: **Hypothesis-driven indices: Dietary patterns: Lifestyle indices: Hyperinsulinaemia: Insulin resistance: C-peptide: TAG: HDL**

Hyperinsulinaemia and insulin resistance are considered important underlying mechanisms linking poor dietary and lifestyle behaviours to the development of multiple chronic diseases and conditions. For example, studies suggest that hyperinsulinaemia is associated with higher risk of colorectal adenomas[1] and colorectal cancer independent of adiposity[2,3], and insulin resistance has been consistently linked to obesity, inflammation, heart disease and type 2 diabetes[4–6]. Although specific dietary factors have been shown to influence insulin resistance and secretion[7,8], dietary patterns or indices that include multiple dietary factors and account for the complex interactions among nutrients and foods may be more predictive of diet–disease associations[9,10]. Other lifestyle factors that have been linked to hyperinsulinaemia and insulin resistance are body weight and physical activity (PA)[11–14]. PA plays an important role in the prevention of insulin

Abbreviations: EDIH, empirical dietary index for hyperinsulinaemia; EDIR, empirical dietary index for insulin resistance; ELIH, empirical lifestyle index for hyperinsulinaemia; ELIR, empirical lifestyle index for insulin resistance; GI, glycaemic index; HPFS, Health Professionals Follow-up Study; NHS, Nurses' Health Study; PA, physical activity.

* **Corresponding author:** F. K. Tabung, email ftabung@hsph.harvard.edu
† Co-first author.

insensitivity[14], whereas increased body weight has a direct association with insulin resistance[11]. Therefore, combining diet, exercise and body weight in a lifestyle index would likely be more predictive of hyperinsulinaemia and insulin resistance than each of these factors considered separately.

At present, the most common dietary index used to assess the ability of diets to stimulate insulin secretion is the glycaemic index (GI). The GI classifies carbohydrate-containing foods by their ability to raise postprandial blood glucose concentration relative to glucose or white bread[15], and therefore indirectly assesses immediate insulin responses to food intake. However, it neglects dietary factors such as proteins and fats that are also important in insulin secretion. Moreover, the GI does not quantify the long-term effects of diet on glycaemia. As an improvement on the GI, our group previously developed a food insulin index to directly quantify postprandial insulin response[16]. However, this index was not predictive of C-peptide concentrations[16]. The lack of predictive ability may be because the insulin index, similar to the GI, assesses postprandial insulin response to the intake of specific foods, and therefore is limited to quantifying short-term insulin response rather than the long-term effects of whole diets on insulinaemia. Hence, we developed dietary and lifestyle patterns that assess the insulinaemic potential of usual diets and lifestyles to reflect long-term insulin exposure and overall insulin resistance, the more relevant exposure for chronic disease prevention.

Previously, our group derived a dietary pattern associated with hyperinsulinaemia and found this pattern to be significantly associated with colorectal cancer risk[17]. However, the sample size used to derive this pattern was small (n 833), and the pattern was applied in the same cohort. The objectives of our current study were 3-fold: first, we updated the previously developed dietary pattern using the currently available larger sample of women and additionally developed separate dietary and lifestyle patterns predictive of hyperinsulinaemia, as well as insulin resistance; second, in validation studies, we evaluated how well these patterns predicted concentrations of insulin response biomarkers in independent samples of men and women; and, third, we examined the joint influence of diet, body weight and PA on clinically relevant hyperinsulinaemia and insulin resistance.

Methods

Study populations

The Nurses' Health Study (NHS), the Nurses' Health Study-II (NHS-II) and the Health Professionals Follow-up Study (HPFS) are ongoing prospective cohorts established in 1976, 1989 and 1986, respectively. The NHS (n 121 701) enrolled female registered nurses aged 30–55 years, whereas the NHS-II (n 116 430) enrolled younger female registered nurses aged 25–42 years[18]. The HPFS (n 51 529) enrolled male health professionals aged 40–75 years. Blood samples were collected from subpopulations of the NHS (n 32 826) in 1989–1990, of the NHS-II (n 29 611) between 1996 and 1999 and of the HPFS (n 18 225) from 1993 to 1994[19]. Blood sample collection was conducted using similar protocols for all cohorts. The procedures including collection, handling and storage have been previously summarised[20]. In the current study, we used data from previous matched case–control studies nested within each of the three cohorts that measured fasting concentrations of plasma C-peptide, TAG and HDL-cholesterol. In the NHS, 5812 women with C-peptide data and 3929 women with data on TAG and HDL-cholesterol were included in the development of the dietary and lifestyle indices. For the validation studies, there were 4002 men with C-peptide data and 3559 men with TAG and HDL data in the HPFS cohort, and there were 1717 women with C-peptide data and 1008 women with TAG and HDL data in the NHS-II cohort. The Institutional Review Boards at Brigham and Women's Hospital and at Harvard T.H. Chan School of Public Health approved this study.

Biomarker assessment

For the current analysis, we utilised fasting plasma C-peptide concentrations to assess hyperinsulinaemia. Compared with insulin, C-peptide has proven to be a better measure of β-cell secretory activity as it is not extracted by the liver, has a slower metabolic clearance rate and does not cross-react with antibodies of insulin[21]. To assess insulin resistance, we utilised the ratio of fasting TAG:fasting HDL-cholesterol, which has been shown to be significantly correlated with insulin resistance[22]. TAG:HDL-cholesterol is also a simple and clinically useful measure to identify apparently healthy individuals who are insulin resistant[23–25].

Procedures for the measurement of fasting plasma insulinaemic markers (C-peptide, TAG and HDL) in the NHS, NHS-II and HPFS have previously been described[26,27]. C-peptide was measured by ELISA (Diagnostic Systems Laboratories/Beckman Coulter). HDL-cholesterol and TAG were measured by standard methods with reagents from Roche Diagnostics and Genzyme[26,27]. The intra-assay CV from blinded quality control samples were <12 % for C-peptide and <1·8 % for TAG and HDL across batches.

In nested case–control studies in which these biomarkers were measured, samples from cases and their matched controls were analysed in the same batch. Quality control samples were randomly interspersed among case–control samples, and laboratory personnel were blinded to quality control and case–control status for all assays. Biomarkers were measured in multiple batches over several years. There may be differences in mean biomarker levels by batch due to different reagents, technicians or laboratories, but also due to differences in the participants in each batch. We therefore used a three-step method, previously described by Rosner *et al.*[28], to re-calibrate biomarker concentrations across several batches to the value of an 'average batch', accounting for true variability across batches, because of different distributions of predictors of the biomarker across batches: (i) we constructed a linear regression model with biomarker levels as the dependent variables and batch indicators as well as variables that may vary by biomarker levels and by batch (regular aspirin/non-steroidal anti-inflammatory drugs (NSAID) use, age at blood sample collection, PA, smoking status, diabetes, other chronic diseases/conditions and case–control status, as well as menopausal

status and postmenopausal hormone use in women) as the independent variables; (ii) next, we calculated the average batch β coefficient by summing the batch indicator β and dividing by the total number of batches; and (iii) finally, we calculated the difference between each batch β and the average β and re-calibrated biomarker concentrations by subtracting this difference from the original biomarker concentration. The re-calibrated biomarkers were then used in the analyses. The correlation between the re-calibrated and the uncalibrated TAG:HDL-cholesterol was 0·96 and 0·85 for C-peptide in the NHS; therefore, we used the uncalibrated TAG:HDL-cholesterol and calibrated C-peptide in the primary analyses and conducted sensitivity analyses with the re-calibrated TAG:HDL-cholesterol and uncalibrated C-peptide.

Assessment of dietary and non-dietary data

Dietary data are updated every 4 years in the NHS (since 1980), the NHS-II (since 1991) and in the HPFS (since 1986) with a validated, semi-quantitative FFQ that assessed diet intake during the previous 1 year[29–31]. We used dietary data from the questionnaires closest to the blood draw – that is, the 1990 FFQ for the NHS, the 1999 FFQ for the NHS-II and the 1994 FFQ for the HPFS. Participants with excessive missing items (≥70) in the FFQ or with implausibly low or high energy intakes (<2510 or >14 644 kJ/d (<600 or >3500 kcal/d) for women and <3347 or >17 573 kJ/d (<800 or >4200 kcal/d) for men) were excluded[32].

All three cohorts collected non-dietary data (e.g. medical history and health practices) and updated the data through biennial, self-administered questionnaires. We calculated participants' BMI (kg/m^2) using height (metres) reported at baseline for each cohort and weight (kg) reported in the questionnaire closest to blood draw. Participants reported their smoking status (never, former, current), and we calculated PA, expressed in metabolic equivalent (MET)-h/week, by summing the average MET-h/week for the following activities: tennis/squash/racquetball, rowing, calisthenics, walking, jogging, running, bicycling and swimming. The reproducibility and validity of the PA questionnaire have been evaluated previously[33,34]. Regular use of aspirin or other NSAID was defined as use of ≥2 standard tablets (325-mg) of aspirin or ≥2 tablets of NSAID/week. We derived a chronic disease co-morbidity score by summing the presence = 1/absence = 0 of the following chronic diseases/conditions: hypercholesterolaemia, cancer, high blood pressure, heart disease and rheumatoid/other arthritis.

Development of the indices of lifestyle and dietary insulinaemic potential

We developed four indices to assess the insulinaemic potential of whole diets and lifestyles: the empirical dietary index for hyperinsulinaemia (EDIH) and the empirical lifestyle index for hyperinsulinaemia (ELIH), which also include BMI and PA as components; the empirical dietary index for insulin resistance (EDIR) and the empirical lifestyle index for insulin resistance (ELIR), which also include BMI and PA as components.

Of the three cohorts, the NHS had the largest sample of participants with biomarker data; therefore, we used dietary, lifestyle and biomarker data (C-peptide, TAG and HDL) in the NHS to develop the indices, and based the scores on food groups rather than on nutrients to approximate how people perceive dietary intake. We first calculated daily intakes per 4184 kJ (1000 kcal) of thirty-nine previously defined food groups[32] from the 1990 FFQ. The grouping scheme was based on the similarity of the nutrient profiles or culinary usage among the foods[32]. We then used four separate stepwise multivariable-adjusted linear regression analyses to identify the most important component food groups and lifestyle factors contributing to hyperinsulinaemia (with C-peptide concentrations as the dependent variable) and to insulin resistance (with TAG:HDL-cholesterol as the dependent variable), with the thirty-nine food groups as independent variables, and a significance level of $P = 0{\cdot}1$ for entry into and retention in the model. BMI and PA were added to the list of the thirty-nine food group predictors in models to develop the lifestyle indices. Intakes of the food groups identified in the stepwise linear regression analyses were weighted by the regression coefficients derived from the final stepwise linear regression model and then summed to constitute the indices. All four index scores assess the insulinaemic potential of diet on a continuum from maximally low insulinaemic potential to maximally high insulinaemic potential, with higher (more positive) scores indicating higher insulinaemic diets or lifestyles (hyperinsulinaemia or insulin resistance) and lower (more negative) scores indicating low insulinaemic or insulin-sensitive diets or lifestyles.

Sensitivity analyses

In the sensitivity analyses, we created three potential alternative versions of both the EDIH and the EDIR by (i) using uncalibrated C-peptide and calibrated TAG:HDL-cholesterol, (ii) using unweighted components, thus assuming that all components contribute equally to the total score, and (iii) by constructing indices only for control subjects of the nested case–control studies (although all the nested case–control studies that generated data for the current study used pre-diagnostic blood samples from chronic disease-free participants).

In addition, we compared the predictive ability of the previously developed C-peptide dietary pattern. This pattern was high in red meat, high-energy beverages, fish and creamy soup intakes and low in coffee, high-fat dairy and wholegrain intakes[17]. Finally, we compared the predictive ability of the EDIH and the EDIR with that of the previously developed insulin index. The insulin index has been described previously; its values compare the postprandial plasma insulin response of a specific food relative to a reference food[16].

Statistical analysis

Where it is not explicitly stated, the analyses described for the EDIH and the EDIR were also applied to their respective lifestyle versions. We described participants' characteristics using mean values and standard deviations for continuous variables or geometric means and CV for log-transformed variables and frequencies (%) for categorical variables. Concentrations of all biomarkers were back-transformed to their original units

(e^x, where x is the transformed biomarker value) because biomarkers were log-transformed using natural logarithms before analyses.

In the NHS, we calculated correlation coefficients between the EDIH or the EDIR, their alternative versions and the insulinaemic markers. We also assessed the distribution of the absolute average concentrations of C-peptide across quintiles of EDIH and TAG:HDL-cholesterol across quintiles of EDIR, stratified by joint categories of BMI and PA as follows: lean and active (BMI < 25 kg/m^2 and PA ≥ median PA), lean and sedentary (BMI < 25 kg/m^2 and PA < median PA), overweight/obese and active (BMI ≥ 25 kg/m^2 and PA ≥ median PA), and overweight/obese and sedentary (BMI ≥ 25 kg/m^2 and PA < median PA). The multivariable models were adjusted for the following covariates: age at blood draw (years, continuous), PA (MET-h/week, continuous), smoking status (never, former, current), regular aspirin/NSAID use (yes/no), case–control status, history of diabetes (yes/no), chronic disease co-morbidity score and additionally for menopausal status and postmenopausal hormone use. BMI was not controlled for in the multivariable models because it has been shown to mediate[35,36] and/or modify[17] the association between diet and insulin markers; thus, controlling for BMI could result in attenuation of true associations or loss of statistical power to detect true associations.

In the validation studies in which we evaluated how well the indices predicted concentrations of insulin response biomarkers in the HPFS and NHS-II samples, we calculated scores for the EDIH and EDIR and their potential alternative versions, as well as estimated correlations among the index scores and biomarkers (C-peptide for hypersinsulinaemia and TAG:HDL-cholesterol for insulin resistance). In addition, we assessed the distribution of the absolute average concentrations of C-peptide across quintiles of EDIH and TAG:HDL-cholesterol across quintiles of EDIR, stratified by joint categories of BMI,PA described above. To determine whether there were clinically relevant differences in the insulinaemic potential of diet between these categories, we used clinically relevant cut-off points – 1·8 ng/ml for C-peptide[37,38] and 3 for TAG:HDL-cholesterol[25,39] (values considered to be the upper limit of normal) to dichotomise the biomarkers. Participants with values ≥1·8 ng/ml were classified as having high C-peptide concentrations, whereas those with TAG:HDL-cholesterol >3 had high TAG:HDL-cholesterol ratio. We then calculated proportions of participants with clinically high levels of biomarkers across dietary index quintiles in each category of BMI,PA.

The associations between the EDIH or EDIR and their respective outcome biomarkers were assessed in multivariable-adjusted linear regression models using relative concentrations of the biomarkers predicted in higher EDIH or EDIR quintiles, with the lowest quintile as the reference (e.g., concentration in quintile 5/concentration in quintile 1). We used the continuous index adjusted for multiple covariates to assess the trend of biomarker concentrations across quintiles of the categorised index. All multivariable models were adjusted for the previously described potential confounding variables.

In sensitivity analyses, we applied each of the three alternative versions of the EDIH or EDIR (scores developed using uncalibrated C-peptide and calibrated TAG:HDL-cholesterol, scores developed using unweighted components, scores developed in control subjects only) in multivariable-adjusted linear regression models to predict relative concentrations of the biomarkers. In addition, we compared the predictive ability of the previously developed C-peptide dietary pattern and the insulin index with that of the EDIH and EDIR. Although participants were free from diabetes at blood sample collection, we excluded participants identified to have diabetes during the nested case–control studies, and compared findings with those from all participants.

All analyses were conducted using SAS version 9.3 for UNIX. All tests were two-sided, and 95% CI not including 1 were considered to indicate statistically significant results.

Results

Of the thirty-nine food groups examined, eighteen were identified as significant contributors to the EDIH, with thirteen of them positively associated and five of them inversely associated with C-peptide concentrations (Table 1). The ELIH had fourteen components: seven components including BMI were positively associated with C-peptide, whereas the remaining seven components including PA were inversely associated with C-peptide concentrations. Common to both dietary and lifestyle hyperinsulinaemia indices were red meat, margarine, creamy soups and butter (positive associations) as well as high-fat dairy products, wine, coffee and whole fruit (inverse associations). The EDIR had eighteen components: ten were positively associated with TAG:HDL-cholesterol, whereas eight were inversely associated with TAG:HDL-cholesterol. The ELIR had seventeen components: eleven including BMI were positively associated with TAG:HDL-cholesterol, whereas the remaining six including PA were inversely associated with TAG:HDL-cholesterol. Common to both dietary and lifestyle insulin-resistance indices were margarine, red meat, refined grains, processed meat, tomatoes, other vegetables and low-energy beverages (positive associations) as well as coffee, wine, high-fat dairy products, liquor and green leafy vegetables (inverse associations) (Table 1). The potential alternative versions were similar and mainly differed from the EDIH and the EDIR in the number of components (online Supplementary Table S1).

In the NHS, the proportion of overweight women in the highest quintile of both the EDIH and the EDIR was approximately 2 times higher than the proportion in the lowest quintile. Similarly, the proportion of lean and active participants was highest in quintile 1 and lowest in quintile 5. The proportion of participants with ≥3 chronic diseases/conditions in the highest quintile was >2 times higher than that in the lowest quintile (Table 2). Both dietary indices showed moderate correlations with biomarkers. For example, the Spearman's correlation coefficient was 0·21 for the EDIH and C-peptide and 0·32 for the EDIR and TAG:HDL-cholesterol. The correlations were stronger for the two lifestyle indices, with correlations coefficients of 0·47 between the ELIH and C-peptide and 0·46 between the ELIR and TAG:HDL-cholesterol (Table 3). In addition, the EDIH and the EDIR were highly correlated with their potential alternative versions, but correlations with the insulin index were low – for example,

Table 1. Components of the indices to assess the insulinaemic potential of diet and lifestyle; the Nurses' Health Study, 1990

Empirical dietary index for hyperinsulinaemia			Empirical lifestyle index for hyperinsulinaemia			Empirical dietary index for insulin resistance			Empirical lifestyle index for insulin resistance		
Food group*	Weight†	R^2‡	Food group*	Weight†	R^2‡	Food group*	Weight†	R^2‡	Food group*	Weight†	R^2‡
Positive associations			Positive associations			Positive associations			Positive associations		
Red meat	0·250	0·008	BMI (kg/m^2)	0·051	0·187	Low-energy beverages	0·116	0·014	BMI (kg/m^2)	0·047	0·151
Low-energy beverages	0·053	0·004	Margarine	0·041	0·001	Margarine	0·121	0·013	Refined grains	0·076	0·003
Cream soups	0·787	0·003	Liquor	0·072	0·001	Red meat	0·328	0·009	Red meat	0·181	0·003
Processed meat	0·199	0·002	Cream soups	0·536	0·001	Refined grains	0·102	0·006	Margarine	0·099	0·003
Margarine	0·054	0·002	Butter	0·058	0·001	processed meats	0·327	0·004	Tomatoes	0·135	0·002
Poultry	0·183	0·002	Red meat	0·089	0·001	Tomatoes	0·145	0·002	Low-energy beverages	0·051	0·002
Butter	0·094	0·001	Fruit juice	0·042	0·001	Other vegetables	0·126	0·001	Fruit juice	0·068	0·001
French fries	0·581	0·001	Inverse associations			Other fish	0·155	0·001	Potatoes	0·160	0·001
Other fish	0·172	0·001	Coffee	−0·020	0·002	Fruit juice	0·052	0·001	Processed meat	0·124	0·001
High-energy beverages	0·104	0·001	Whole fruit	−0·029	0·002	Creamy soups	0·519	0·001	Other vegetables	0·070	0·001
Tomatoes	0·095	0·001	Wine	−0·071	0·002	Inverse associations			Tea	0·027	0·001
Low-fat dairy products	0·025	0·001	Physical activity (MET-h/week)	−0·001	0·001	Coffee	−0·070	0·018	Inverse associations		
Eggs	0·124	0·001	High-fat dairy products	−0·054	0·001	Wine	−0·261	0·011	Coffee	−0·041	0·007
Inverse associations			Snacks	−0·024	0·001	Liquor	−0·204	0·006	Wine	−0·171	0·004
Wine	−0·165	0·009	Salad dressing	−0·059	0·001	Beer	−0·210	0·002	Liquor	−0·122	0·002
Coffee	−0·035	0·005				Green leafy vegetables	−0·076	0·001	High-fat dairy products	−0·064	0·001
Whole fruits	−0·029	0·003				High-fat dairy products	−0·066	0·001	Physical activity (MET-h/week)	−0·001	0·001
High-fat dairy products	−0·046	0·001				Dark yellow vegetables	−0·103	0·001	Green leafy vegetables	−0·064	0·001
Green leafy vegetables	−0·055	0·001				Nuts	−0·078	0·001			

* The food groups (servings/d) retained were defined as follows: red meats (4–6 oz beef, 4–6 oz pork, 4–6 oz lamb, 1 patty hamburger); processed meat (1 piece or 1 slice processed meat, 2 slices bacon, 1 hot dog); low-energy beverages (1 glass, 1 bottle or 1 can of low-energy cola, other low-energy carbonated beverages); cream soups (1 cup chowder or cream soup); 1 pat margarine; poultry (4–6 oz chicken or turkey with or without skin); high-energy beverages (1 glass, 1 bottle or 1 can of cola with sugar, other carbonated beverages with sugar, fruit punch drinks); 1 pat butter; 4-oz French fries; other fish (3–5 oz canned tuna, shrimp, lobster, scallops, fish and other seafood other than dark meat fish); low-fat dairy products (8-oz glass skimmed or low-fat milk, 1/2 cup sherbet or ice milk, 1 cup yogurt); tomatoes (1 fresh tomato, 1 small glass of tomato juice, 1/2 cup of tomato sauce); 1 egg; cruciferous vegetables (1/2 cup of broccoli, coleslaw and uncooked cabbage; cooked cabbage; cauliflower; Brussels sprouts; kale, mustard and chard greens; sauerkraut); wine (4-oz glass of red wine, white wine); 1 cup coffee; high-fat dairy products (8-oz glass whole milk, cream, 1 tablespoon sour cream, 1/2 cup ice cream, 1 oz cream cheese, 1 oz or 1 slice other cheese); green leafy vegetables (1/2 cup spinach, serving of iceberg or head lettuce, serving of romaine or leaf lettuce); whole fruit (1-oz or small-pack raisins, 1/2 grapes, 1 avocado, 1 banana, 1/4 melon cantaloupe, 1 slice watermelon, 1 fresh apple or pear or 1/2 cup canned, 1 orange, 1/2 grapefruit, 1/2 cup strawberries, 1/2 cup blueberries, 1 fresh or 1/2 canned peaches, 1 fresh or 1/2 canned apricots or plums); dark yellow vegetables (1/2 cup carrots, 1/2 cup yellow (winter) squash, 1/2 cup yams, 1/2 cup sweet potatoes); snacks (1 small bag or 1-oz potato chips, corn chips or popcorn, 1 crackers); 1 pat butter; fruit juice (1 small glass of apple juice or cider, orange juice, grapefruit juice, other fruit juice); liquor (1 drink or 1 shot whiskey, gin, etc.); salad dressing (1 tablespoon oil and vinegar salad dressing); refined grains (1 slice white bread, 1 English muffins, 1 bagel or roll, 1 muffin or biscuit, 1 cup white rice, 1 cup pasta, 1 serving of pancakes or waffles); beer (1 glass, 1 bottle or 1 can); 1 cup coffee; 1 cup tea (not herbal tea); potatoes (1 baked or boiled, 1 cup mashed); other vegetables (4-inch stick celery, 1 fresh, cooked or canned mushroom, 1/2 green pepper, 1 ear or 1/2 cup frozen or canned corn, 1/2 cup mixed vegetables, eggplant, 1/2 cup zucchini, 1/2 alfalfa sprouts, 1/4 cucumber); whole grains (1 cup cooked oatmeal, 1 cup other cooked breakfast cereal, 1 slice dark bread, 1 cup brown rice, 1 cup other grains, 1 tablespoon bran added to food, 1 tablespoon wheat germ).

† Weights are regression coefficients derived from the final step of the stepwise linear regression models. Each weight represents the contribution of the corresponding index component to the total weighted index score.

‡ The partial R^2 represents the proportion of variance in biomarkers explained by the index component.

Table 2. Participant characteristics in quintiles (Q) of the insulin response dietary patterns; the Nurses' Health Study, 1990 (Mean values and standard deviations; numbers and percentages)

	Empirical dietary index for hyperinsulinaemia (*n* 5812)						Empirical dietary index for insulin resistance (*n* 3929)					
	Q1 (*n* 1162) (–0·57 to <0·09)		Q3 (*n* 1162) (0·16 to <0·21)		Q5 (*n* 1162) (0·28 to 0·75)		Q1 (*n* 785) (–1·15 to <0·06)		Q3 (*n* 786) (0·19 to <0·28)		Q5 (*n* 786) (0·39 to 1·25)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Fasting C-peptide (ng/ml)*	1·8	0·9	2·2	0·9	2·6	0·9	NA	NA	NA	NA	NA	NA
Fasting TAG:HDL-cholesterol*	NA	NA	NA	NA	NA	NA	1·62	1·0	2·29	1·0	3·19	1·1
Insulin index	42·5	5·8	43·0	4·8	41·8	4·3	40·4	6·3	43·4	4·3	42·9	4·1
Age (years)	59·3	6·5	58·2	6·9	56·2	7·4	59·2	6·1	59·6	6·3	58·8	6·7
Physical activity (MET-h/week)	18·6	18·7	17·5	30·5	12·9	21·4	17·4	18·7	15·2	19·7	13·5	18·4
Alcohol (servings/d)†	0·7	1·0	0·4	0·7	0·3	0·7	1·1	1·3	0·3	0·5	0·1	0·3
BMI (kg/m^2)	24·0	3·6	25·7	4·3	27·7	5·6	24·3	3·9	26·3	5·1	29·3	6·0
	n	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%
Overweight/obese (≥25 kg/m^2)	373	32·1	570	49·1	728	62·7	279	35·5	396	50·4	578	73·5
Current smokers	125	10·8	116	10·0	146	12·6	247	31·5	165	21·0	123	15·7
Regular aspirin/NSAID users	393	33·8	436	37·5	484	41·7	277	35·3	289	36·8	321	40·8
BMI–physical activity combinations‡												
Lean and active	479	41·2	334	28·7	203	17·5	281	35·8	198	25·2	102	13·0
Lean and sedentary	310	26·7	258	22·2	231	19·9	225	28·7	192	24·4	106	13·5
Overweight/obese and active	211	18·2	272	23·4	259	22·3	165	21·0	188	23·9	229	29·1
Overweight/obese and sedentary	162	13·9	298	25·7	469	40·4	114	14·5	208	26·5	349	44·4
Chronic disease/conditions co-morbidity score§												
No chronic disease/condition	517	44·5	521	44·8	517	44·5	356	45·4	283	36·0	218	27·7
1 chronic disease/condition	415	35·7	376	32·4	374	32·2	261	33·3	260	33·1	238	30·3
2 chronic diseases/conditions	184	15·8	183	15·8	199	17·1	116	14·8	173	22·0	191	24·3
≥3 chronic diseases/conditions	46	4·0	82	7·1	72	6·2	52	6·6	70	8·9	139	17·7
Diabetes (yes)	6	0·5	17	1·5	23	2·0	46	5·9	108	13·7	259	33·0
Postmenopausal women	1024	88·1	974	83·8	889	76·5	726	92·5	714	90·8	697	88·7
Postmenopausal hormone user	679	58·4	633	54·5	535	46·0	490	62·4	455	57·9	424	53·9

NSAID, non-steroidal anti-inflammatory drugs.

* Geometric means (CV) are presented for the biomarkers (fasting plasma samples) because all biomarkers were log-transformed before analyses; the Quan-Zhang formula; $CV = (e^{SD} - 1)^{1/2(40)}$ was used to calculate CV.

† Alcohol intake was the sum of wine (4 oz glass), beer (1 bottle, can or glass) or liquor (1 drink or shot) intakes.

‡ Categories of BMI and physical activity (PA) combinations were created as follows: lean and active (BMI < 25 kg/m^2 and PA ≥ median PA), lean and sedentary (BMI < 25 kg/m^2 and PA < median PA), overweight/obese and active (BMI ≥ 25 kg/m^2 and PA ≥ median PA) and overweight/obese and sedentary (BMI ≥ 25 kg/m^2 and PA < median PA). Median PA = 10·2 MET-h/week for women with C-peptide data and 9·10 MET-h/week for those with TAG:HDL-cholesterol data.

§ Chronic diseases/conditions included in the score were hypercholesterolaemia, cancer, diabetes, high blood pressure, heart disease and rheumatoid/other arthritis.

although the EDIH had a correlation coefficient of 0·90 with the version developed in control subjects, its correlations with the insulin index was –0·07. Corresponding correlations for the EDIR were 0·89 and 0·14, respectively (online Supplementary Table S2).

In multivariable-adjusted models in the NHS, the EDIH and the EDIR were significantly associated with C-peptide and TAG:HDL-cholesterol. The C-peptide concentration of women in the highest quintile of the EDIH was 40 % (95 % CI 34, 46 %; $P_{trend} < 0·0001$) higher than that of women in the lowest quintile. Similarly, women in the highest quintile of the EDIR had a 67 % (95 % CI 55, 80 %; $P < 0·0001$) higher concentration of TAG:HDL-cholesterol than women in the lowest quintile. The corresponding contrasts for the ELIH and the ELIR were 97 % (95 % CI 89, 106 %) and 127 % (95 % CI 111, 145 %), respectively (Table 4). Multivariable-adjusted analyses excluding women with diabetes were not materially different (online Supplementary Table S3). In stratified analyses, there were large differences in C-peptide concentrations in EDIH quintiles across combinations of BMI,PA. Women in the overweight/obese and sedentary categories had the highest concentrations of C-peptide, whereas those in the lean, active category had the lowest concentrations. In addition, there were significant trends of increasing TAG:HDL-cholesterol concentrations within joint strata of BMI and PA (Fig. 1).

In the validation studies using HPFS and NHS-II data, we observed similar trends in participant characteristics as in the NHS. Concentrations of C-peptide and TAG:HDL-cholesterol increased monotonically across quintiles of their respective dietary and lifestyle indices. For example, between extreme index quintiles in the HPFS, there was a 25 and 82 % increase in C-peptide for the EDIH and the ELIH respectively, and a 60 and 132 % increase in TAG:HDL-cholesterol for the EDIR and the ELIR, respectively (online Supplementary Table S4 for the EDIH and the EDIR and online Supplementary Table S5 for the ELIH and the ELIR). Moreover, we found similar correlation patterns for the indices and biomarkers in the HPFS and NHS-II samples as in the NHS – that is, moderate correlations between dietary indices and biomarkers, stronger correlations between lifestyle indices and biomarkers (Table 3) and very strong correlations between dietary indices and potential alternative versions, but low-to-moderate correlations with the insulin index and the

Table 3. Spearman's correlations coefficients among the insulinaemic dietary and lifestyle patterns and fasting plasma biomarker concentrations in the three cohorts

	C-peptide		
Empirical dietary indices for hyperinsulinaemia	NHS	NHS-II	HPFS
C-peptide	1	1	1
EDIH	0·21	0·20	0·14
ELIH	0·47	0·43	0·36
Unweighted EDIH	0·16	0·16	0·09
Unweighted ELIH	0·28	0·24	0·19
EDIH in controls	0·20	0·19	0·14
EDIH with unadjusted C-peptide	0·20	0·21	0·13
Previously developed C-peptide dietary pattern	0·11	0·12	0·09
Insulin index	−0·03	−0·03*	−0·06
	TAG:HDL-cholesterol		
Empirical dietary indices for insulin resistance	NHS	NHS-II	HPFS
TAG:HDL-cholesterol	1	1	1
EDIR	0·32	0·16	0·21
ELIR	0·46	0·35	0·39
Unweighted EDIR	0·28	0·10	0·19
Unweighted ELIR	0·27	0·24	0·16
EDIR in controls	0·28	0·16	0·18
EDIR with adjusted TAG, HDL	0·31	0·18	0·21
Insulin index	0·06	0·07	0·05

NHS, Nurses' Health Study, 1990; NHS-II, Nurses' Health Study-II, 1999; HPFS, Health Professional Follow-up Study, 1994; EDIH, empirical dietary index for hyperinsulinaemia; ELIH, empirical lifestyle index for hyperinsulinaemia; EDIR, empirical dietary index for insulin resistance; ELIR, empirical lifestyle index for insulin resistance.
* $P > 0·05$.

previously developed C-peptide dietary pattern (online Supplementary Table S2). The insulin index was inversely correlated with C-peptide and with the EDIH. In the HPFS, the correlation between the EDIH and the EDIR was 0·63.

All four indices were significantly associated with their respective biomarkers in HPFS and NHS-II, with stronger associations observed for the two lifestyle indices than their diet-only counterparts (Table 4). For example, in the HPFS, the relative concentration of C-peptide was 29 % (95 % CI; 22 %, 27 %; $P_{trend} < 0·0001$) higher in the highest quintile of the EDIH compared with the lowest quintile, whereas the concentration of TAG:HDL-cholesterol was 44 % (95 % CI; 34 %, 55 %; $P_{trend} < 0·0001$) higher in quintile 5 of the EDIR compared with quintile 1. Corresponding associations for the lifestyle indices were as follows: 78 % (95 % CI 68, 88 %; $P < 0·0001$ for the ELIH and 103 % (95 % CI 89, 119 %; $P_{trend} < 0·0001$) for the ELIR (Table 4). Excluding participants with diabetes did not materially change these findings (online Supplementary Table S3). In the HPFS, there were differences in concentrations of C-peptide and TAG:HDL-cholesterol across index quintiles and in categories of BMI,PA, with overweight/obese and sedentary men having the highest biomarker levels compared with overweight/obese and active men or lean, active or sedentary men (Fig. 2). The proportion of participants with clinically high C-peptide concentrations across each EDIH quintile was 1·5–2 times higher among overweight/obese and sedentary men than among lean and active men, whereas the proportion with high TAG:HDL-cholesterol levels was 2 to 3 times higher across each EDIR quintile among overweight/obese and sedentary men than among lean and active men. Among men classified as lean and active, a higher proportion of those consuming diets with high insulinaemic potential had clinically high biomarker levels than those consuming insulin-sensitive diets (Fig. 3).

Results from the sensitivity analyses in both men and women showed that the associations between dietary patterns developed only in control subjects and those with uncalibrated C-peptide and uncalibrated TAG:HDL-cholesterol with biomarkers were reasonably similar to the associations obtained with the EDIH or the EDIR. However, associations for the unweighted versions and the previously developed C-peptide pattern were smaller in magnitude. In contrast, the insulin index was not predictive of C-peptide concentrations in both men and women. The relative concentrations were as follows: 0·94 (95 % CI 0·89, 1·00; $P_{trend} = 0·03$) for men and 0·99 (95 % CI 0·91, 1·09; $P_{trend} < 0·90$) for women, comparing extreme index quintiles, although there was a trend towards an inverse association in men. The insulin index, however, had a direct (but smaller compared with the EDIR) association with TAG:HDL-cholesterol in men, 1·20 (95 % CI 1·11, 1·29; $P_{trend} < 0·0001$), but not in women, 1·12 (95 % CI 0·99, 1·26; $P_{trend} = 0·06$), comparing extreme index quintiles. The previously developed C-peptide dietary pattern also had direct associations (although smaller in magnitude) with C-peptide concentrations in both men and women (online Supplementary Table S6).

Discussion

We developed two dietary and two lifestyle indices in a large cohort of women and evaluated their validity in two large independent cohorts of men and women. In all cohorts, the indices were predictive of both the absolute and the relative concentrations of the insulin response biomarkers, although the lifestyle indices were more predictive than the dietary indices. When we applied cut-off points that have been shown to discriminate between clinically high and low biomarker concentrations in adults, we found a consistently higher proportion of participants with high biomarker concentrations across index quintiles within subgroups defined by joint categories of BMI, PA and across BMI,PA categories within index quintiles. These dietary indices assess the long-term insulinaemic potential of whole diets, which is in contrast to the assessment of the acute postprandial glycaemic or insulinaemic potential of specific foods, as has been carried out previously. In addition, the use of the TAG:HDL-cholesterol ratio to derive the insulin-resistance dietary pattern is novel. Although our group previously used C-peptide concentrations to derive a hyperinsulinaemia dietary pattern[17], in the current study, we updated and strengthened this pattern by validating it in two independent cohorts of men and women. Several sensitivity analyses supported the robustness of the EDIH and the EDIR.

The dietary patterns, although empirical, align well with current knowledge. In concordance with the inverse associations found for whole fruits, green leafy vegetables and coffee with hyperinsulinaemia, other studies have shown that higher intakes of coffee as well as a plant-based diet that is high in

Table 4. Adjusted* relative concentrations† of biomarkers in quintiles of insulinaemic dietary and lifestyle patterns in the three cohorts (Relative concentration and 95 % confidence intervals)

		Quintile 2		Quintile 3		Quintile 4		Quintile 5		
	Quintile 1 (Ref.)	Relative concentration	95 % CI	Relative concentration	95 % CI	Relative concentration	95 % CI	Relative concentration	95 % CI	P_{trend}‡
Empirical dietary index for hyperinsulinaemia										
C-peptide (NHS, *n* 5812)										
Age-adjusted	1	1·09	1·04, 1·15	1·20	1·14, 1·25	1·28	1·23, 1·35	1·44	1·38, 1·51	<0·0001
Multivariable-adjusted	1	1·09	1·04, 1·14	1·18	1·13, 1·24	1·27	1·21, 1·32	1·40	1·34, 1·46	<0·0001
C-peptide (HPFS, *n* 4002)										
Age-adjusted	1	1·13	1·07, 1·20	1·16	1·09, 1·23	1·21	1·14, 1·29	1·29	1·22, 1·37	<0·0001
Multivariable-adjusted	1	1·12	1·06, 1·19	1·16	1·09, 1·23	1·22	1·15, 1·29	1·29	1·22, 1·37	<0·0001
C-peptide (NHS-II, *n* 1717)										
Age-adjusted	1	1·05	0·96, 1·15	1·15	1·05, 1·26	1·19	1·09, 1·30	1·37	1·25, 1·50	<0·0001
Multivariable-adjusted	1	1·05	0·96, 1·15	1·13	1·04, 1·24	1·16	1·06, 1·27	1·32	1·21, 1·45	<0·0001
Empirical lifestyle index for hyperinsulinaemia										
C-peptide (NHS, *n* 5812)										
Age-adjusted	1	1·10	1·06, 1·15	1·26	1·21, 1·32	1·54	1·48, 1·60	2·04	1·96, 2·13	<0·0001
Multivariable-adjusted	1	1·10	1·05, 1·14	1·25	1·19, 1·30	1·51	1·44, 1·57	1·97	1·89, 2·06	<0·0001
C-peptide (HPFS, *n* 4002)										
Age-adjusted	1	1·20	1·13, 1·27	1·31	1·24, 1·38	1·46	1·38, 1·54	1·83	1·73, 1·94	<0·0001
Multivariable-adjusted	1	1·19	1·12, 1·25	1·29	1·22, 1·36	1·43	1·35, 1·51	1·78	1·68, 1·88	<0·0001
C-peptide (NHS-II, *n* 1717)										
Age-adjusted	1	1·16	1·06, 1·26	1·31	1·21, 1·43	1·41	1·29, 1·54	1·96	1·80, 2·14	<0·0001
Multivariable-adjusted	1	1·16	1·06, 1·26	1·31	1·21, 1·43	1·41	1·29, 1·54	1·90	1·74, 2·08	<0·0001
Empirical dietary index for insulin resistance										
TAG:HDL-cholesterol (NHS, *n* 3929)										
Age-adjusted	1	1·19	1·11, 1·28	1·41	1·32, 1·52	1·56	1·45, 1·67	1·98	1·84, 2·12	<0·0001
Multivariable-adjusted	1	1·18	1·10, 1·26	1·34	1·25, 1·44	1·43	1·33, 1·54	1·67	1·55, 1·80	<0·0001
TAG:HDL-cholesterol (HPFS, *n* 3559)										
Age-adjusted	1	1·09	1·01, 1·18	1·24	1·15, 1·34	1·35	1·25, 1·45	1·59	1·48, 1·72	<0·0001
Multivariable-adjusted	1	1·11	1·03, 1·19	1·21	1·13, 1·30	1·29	1·20, 1·39	1·44	1·34, 1·55	<0·0001
TAG:HDL-cholesterol (NHS-II, *n* 1008)										
Age-adjusted	1	1·02	0·90, 1·14	1·12	0·99, 1·26	1·32	1·17, 1·49	1·32	1·17, 1·49	<0·0001
Multivariable-adjusted	1	0·98	0·87, 1·10	1·08	0·96, 1·22	1·23	1·09, 1·38	1·19	1·05, 1·34	0·001
Empirical lifestyle index for insulin resistance										
TAG:HDL-cholesterol (NHS, *n* 3929)										
Age-adjusted	1	1·24	1·15, 1·32	1·57	1·46, 1·68	1·98	1·85, 2·12	2·60	2·43, 2·78	<0·0001
Multivariable-adjusted	1	1·24	1·16, 1·33	1·54	1·44, 1·65	1·84	1·72, 1·98	2·27	2·11, 2·45	<0·0001
TAG:HDL-cholesterol (HPFS, *n* 3559)										
Age-adjusted	1	1·28	1·19, 1·37	1·58	1·47, 1·70	1·92	1·79, 2·06	2·34	2·18, 2·52	<0·0001
Multivariable-adjusted	1	1·23	1·15, 1·32	1·49	1·39, 1·60	1·76	1·64, 1·89	2·03	1·89, 2·19	<0·0001
TAG:HDL-cholesterol (NHS-II, *n* 1008)										
Age-adjusted	1	1·12	1·00, 1·26	1·29	1·15, 1·44	1·64	1·46, 1·83	1·90	1·69, 2·13	<0·0001
Multivariable-adjusted	1	1·06	0·95, 1·19	1·21	1·08, 1·36	1·49	1·32, 1·68	1·67	1·48, 1·89	<0·0001

Ref., referent values; NHS, Nurses' Health Study, 1990; NHS-II, Nurses' Health Study-II, 1999; HPFS, Health Professional Follow-up Study, 1994.

* Multivariable-adjusted models were adjusted for regular aspirin/non-steroidal anti-inflammatory drugs use, age, physical activity, smoking status, diabetes, other chronic diseases/conditions and case–control status; NHS and NHS-II models were additionally adjusted for menopausal status and postmenopausal hormone use.

† Values are relative concentrations of fasting plasma biomarkers (i.e. ratio of concentration in higher index quintiles to concentration in the lowest quintile 1 as reference). All values were back-transformed (e^x, where x is the transformed biomarker value) as all biomarkers were transformed using natural log before analyses.

‡ The *P*-value for trend was the *P*-value of the index as a continuous variable, adjusted for all covariates listed in footnote*.

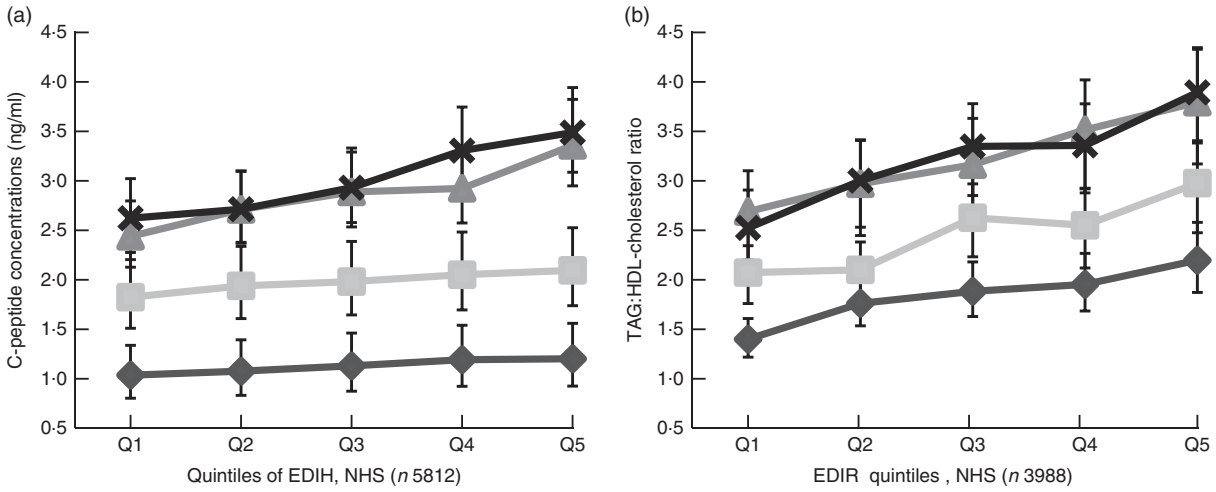

Fig. 1. Multivariable-adjusted biomarker concentrations across quintiles (Q) of (a) the empirical dietary index for hyperinsulinaemia (EDIH) and (b) the empirical dietary index for insulin resistance (EDIR), stratified by joint categories of BMI and physical activity (PA) in the Nurses' Health Study (NHS), 1990. Values are back-transformed (e^x, where x is the transformed biomarker value) predicted mean fasting plasma biomarker concentrations, obtained from linear regression models, adjusted for regular aspirin/non-steroidal anti-inflammatory drugs (NSAID) use, age at blood draw, smoking status, PA, menopausal status, postmenopausal hormone use, diabetes, other chronic diseases/conditions and case–control status. The *P*-value for trend was the *P*-value of the dietary index as a continuous index variable adjusted for all covariates. Categories of BMI and PA combinations were created as follows: lean and active (lean,act; BMI < 25 kg/m^2 and PA ≥ median PA), lean and sedentary (lean,sed; BMI < 25 kg/m^2 and PA < median PA), overweight/obese and active (owt/ob,act; BMI ≥ 25 kg/m^2 and PA ≥ median PA) and overweight/obese and sedentary (owt/ob,sed; BMI ≥ 25 kg/m^2 and PA < median PA). Median PA = 10·2 MET-h/week for women with C-peptide data and 9·10 MET-h/week for those with TAG:HDL-cholesterol data. a: ◆, Lean,act (P_{trend} < 0·0001); ▬, lean,sed (P_{trend} < 0·0002); ▲, owt/ob,act (P_{trend} < 0·0001); ✖, owt/ob,sed (P_{trend} < 0·0001); b: ◆, Lean,act (P_{trend} < 0·0001); ▬, lean,sed (P_{trend} < 0·0001); ▲, owt/ob,act (P_{trend} < 0·0001); ✖, owt/ob,sed (P_{trend} < 0·0001).

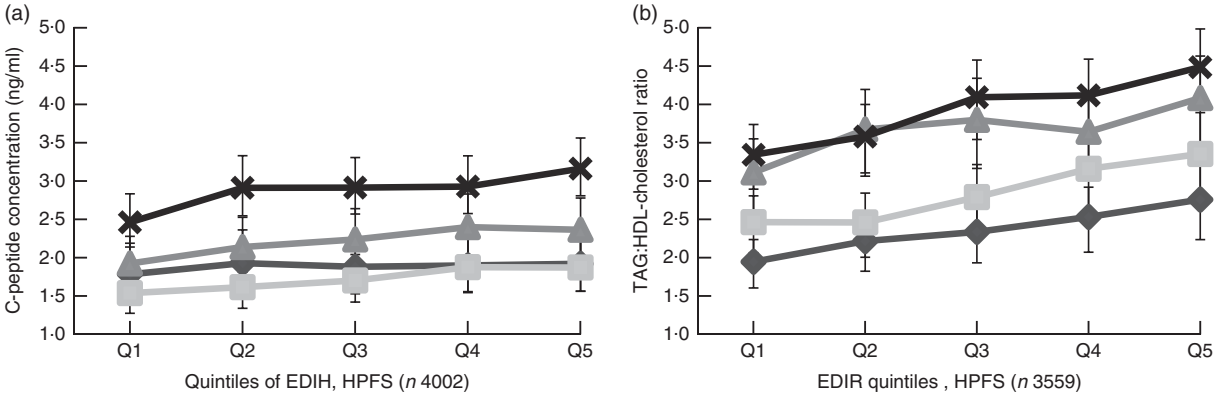

Fig. 2. Multivariable-adjusted biomarker concentrations across quintiles (Q) of (a) the empirical dietary index for hyperinsulinaemia (EDIH) and (b) the empirical dietary index for insulin resistance (EDIR), stratified by joint categories of BMI and physical activity (PA) in the Health Professional Follow-up Study (HPFS), 1994. Values are back-transformed (e^x , where x is the transformed biomarker value) predicted mean fasting plasma biomarker concentrations, obtained from linear regression models, adjusted for regular aspirin/non-steroidal anti-inflammatory drugs (NSAID) use, age, smoking status, PA, diabetes, other chronic diseases/conditions and case–control status. The *P*-value for trend was the *P*-value of the dietary index as a continuous index variable adjusted for all covariates. Categories of BMI and PA combinations were created as follows: lean and active (lean,act; BMI < 25 kg/m^2 and PA ≥ median PA), lean and sedentary (lean,sed; BMI < 25 kg/m^2 and PA < median PA), overweight/obese and active (owt/ob,act; BMI ≥ 25 kg/m^2 and PA ≥ median PA) and overweight/obese and sedentary (owt/ob,sed; BMI ≥ 25 kg/m^2 and PA < median PA). Median PA = 28·1 MET-h/week for men with C-peptide data and 24·8 MET-h/week for men with TAG:HDL-cholesterol data. a: ◆, Lean,act (P_{trend} < 0·09); ▬, lean,sed (P_{trend} < 0·0002); ▲, owt/ob,act (P_{trend} < 0·001); ✖, owt/ob,sed (P_{trend} < 0·0001); b: ◆, Lean,act (P_{trend} < 0·0001); ▬, lean,sed (P_{trend} < 0·0001); ▲, owt/ob,act (P_{trend} < 0·001); ✖, owt/ob,sed (P_{trend} < 0·0001).

fibre, fruits and wholegrains are associated with lower concentrations of C-peptide[7,8,41]. The dietary pattern predictive of insulin resistance is simultaneously influenced by factors that affect both TAG and HDL-cholesterol. We found margarine, refined grains, processed meats, creamy soups and fruit juice to be positively associated with insulin resistance, whereas nuts, alcohol and green leafy vegetables were inversely associated with insulin resistance. Similarly, in previous studies, diets consisting of refined carbohydrates and sweeteners as well as large amounts of SFA and trans fats (as in many cream-based sauces) have been associated with higher TAG concentrations, whereas higher intake of *n*-3-fats such as in nuts and the moderate use of alcohol have been linked to higher levels of HDL-cholesterol[7,42].

We found clinically relevant differences in biomarker concentrations both across dietary index quintiles and across BMI, PA categories. For example, 73 % of overweight/obese and sedentary men consuming the most pro-insulinaemic diets had

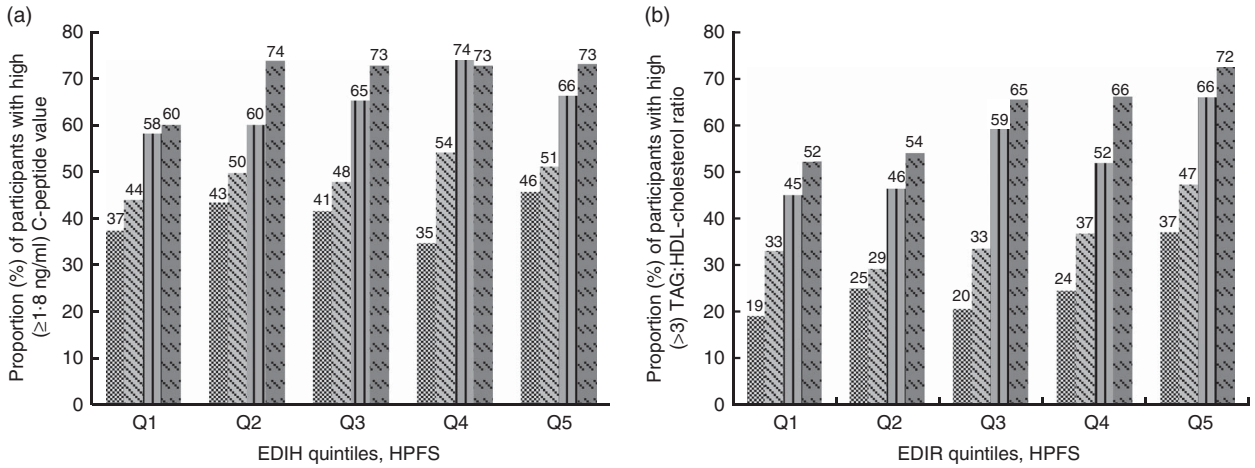

Fig. 3. Distribution of participants (%) with clinically high levels of biomarkers in quintiles (Q) of dietary indices and in joint categories of BMI/physical activity (PA) combinations in the Health Professionals Follow-up Study (HPFS), 1994. Categories of BMI and PA combinations were created as follows: lean and active (lean,act; BMI < 25 kg/m^2 and PA ≥ median PA), lean and sedentary (lean,sed; BMI < 25 kg/m^2 and PA < median PA), overweight/obese and active (owt/ob,act; BMI ≥ 25 kg/m^2 and PA ≥ median PA) and overweight/obese and sedentary (owt/ob,sed; BMI ≥ 25 kg/m^2 and PA < median PA). Median PA = 28·1 MET-h/week for men with C-peptide data and 24·8 MET-h/week for men with TAG:HDL-cholesterol data. a: ▩, Lean,act (*n* 965); ▧, lean,sed (*n* 775); ▮, owt/ob,active (*n* 1038); ▩, owt/ob,sed (*n* 1224); b: ▩, Lean,act (*n* 746); ▧, lean,sed (*n* 660); ▮, owt/ob,active (*n* 830); ▩, owt/ob,sed (*n* 1166).

high C-peptide concentrations (≥1·8 ng/ml) compared with only 37 % of lean and active men consuming the least pro-insulinaemic diets. In addition, 72 % of overweight/obese and sedentary men consuming the most insulin-resistant diets had high TAG:HDL-cholesterol levels (>3) compared with only 19 % of lean and active men consuming the most insulin-sensitive diets. These differences further strengthen the idea that these dietary indices can be useful in identifying populations at risk of hyperinsulinaemia or insulin resistance. Our approach to create lifestyle indices (ELIH and ELIR) is complementary to the stratification of the diet-only indices (EDIH and EDIR) by BMI,PA combinations. Lifestyle indices assess the joint influence of diet, body weight and PA on hyperinsulinaemia and insulin resistance, which is important for public health interventions. The indices assess the insulinaemic potential of diet/lifestyle on a continuum from maximally low insulinaemic to maximally high insulinaemic potential with no optimal cut-off point for classifying individuals as absolutely high or low. Stratifying the diet-only indices by BMI,PA combinations according to established clinically relevant biomarker cut-off points provides further insight on subgroups to target with specific dietary and/or lifestyle interventions to reduce hyperinsulinaemia and/or insulin resistance.

Differences between participants with clinically high and low biomarker levels within quintiles of the dietary indices were observed, despite the low-to-moderate correlations between the indices and the biomarkers. In previous studies, hypothesis-driven dietary patterns have shown low-to-moderate correlations with the biomarkers used to derive the patterns; however, these dietary patterns have shown robust associations with disease risk in independent populations[43,44]. For example, Fung *et al.*[17] reported a correlation coefficient of 0·23 between the dietary pattern predictive of C-peptide and the C-peptide concentrations in NHS, although the pattern showed a significant positive association with colon cancer risk. In addition, a dietary inflammatory index showed low correlations with inflammatory markers, yet strong associations with chronic diseases including cancer[45–47]. This suggests that correlations with biomarkers may not be a direct assessment of the performance of the dietary pattern in disease prediction or clinical significance. For example, among lean and active men, comparing the highest quintile of EDIR to the lowest, the prevalence of clinically high TAG:HDL-cholesterol levels can potentially be reduced by >50 % through diet interventions, even though the EDIR had a low correlation (r 0·15) with TAG:HDL-cholesterol. A low/moderate correlation may also be due to the dietary patterns not capturing other lifestyle behaviours that are associated with the biomarker. Interestingly, when lifestyle factors such as BMI and PA were included, the correlations between the lifestyle indices and the biomarkers were >2 higher than that between the diet-only indices and the biomarkers.

Our group previously created the dietary insulin index to quantify the short-term (postprandial) insulin-secreting ability of specific foods[16]. This index was associated with higher TAG and lower HDL levels, with an indicative inverse association with C-peptide concentrations[16]. In the current study, we compared the predictive ability of the four indices with the insulin index in sensitivity analyses. The insulin index was directly associated with TAG:HDL-cholesterol, which is expected in the context of prevalent insulin resistance, but the correlation was much lower than that of our empirical indices with TAG:HDL-cholesterol. Moreover, the index also showed an inverse trend of association with C-peptide concentrations, which at first seemed counterintuitive but may be understood in the context of our cross-sectional study design using fasting plasma samples – for example, in participants who may usually be consuming a high EDIH/high GI diet; such a diet will elicit higher insulin secretion to reduce the acute postprandial glycaemia. The lowered glucose level will down-regulate

further insulin secretion[48], and blood drawn a couple of hours into the fasting period will therefore show an inverse association (temporarily) between the insulin index (postprandial insulinaemia) and insulin secretion (C-peptide concentration), which may not persist longitudinally.

Our study is not without limitations. We only had one measurement of the insulin markers, which may underestimate validity assessed by correlation coefficients[49]. Given that food intake was self-reported, some measurement error is inevitable, although the validation data showed reasonably good correlations between FFQ and diet records, suggesting that dietary intake is generally well measured in our cohorts[29–31]. The composition of food groups may not be uniform across studies, which would limit the ability to apply the indices across studies in a standardised manner, although investigators may be able to create unified food groups in pooled analyses of primary data or in multi-centre studies, and thus enhance the usefulness of these hypothesis-driven dietary patterns in large-scale epidemiological studies. Study participants in all three cohorts are mostly Caucasian health professionals, but the distributions of most participant characteristics in the three cohorts are generally similar to that of the larger US multi-racial/ethnic population. It is important, however, to further apply the indices in multi-racial/ethnic populations. Other lifestyle factors include smoking and exogenous hormone use, but we focused mainly on BMI and PA in the lifestyle indices because these have been shown to be strongly associated with circulating insulin markers[11–14]. We adjusted for a large number of potential confounding variables including a history of diabetes and other chronic diseases/conditions, but these variables were self-reported, thus allowing the possibility of residual confounding. However, results from the age-adjusted and multivariable-adjusted models were very similar in all cohorts, suggesting that any confounding would have been very minimal.

Conclusion

These novel hypothesis-driven empirically derived dietary and lifestyle indices assess dietary and lifestyle quality on the basis of insulinaemic potential. Their robust associations with the insulin response biomarkers in independent samples suggest their usefulness in assessing the ability of whole diets and lifestyles to stimulate and/or sustain insulin secretion. These indices can be useful in identifying populations at high risk for hyperinsulinaemia or insulin resistance. In addition, the indices may be calculated in a standardised and reproducible manner across different populations, thus circumventing a major limitation of dietary patterns derived from the same study in which they are applied. Moreover, studies without insulin markers data may calculate the index scores to investigate associations between dietary and lifestyle insulinaemic potential and disease outcomes.

Acknowledgements

Dr Jorge E. Chavarro and Dr Frank B. Hu were supported by National Institutes of Health (NIH) grants P30DK046200 and U54 CA155426. The HPFS, NHS and NHS-II cohorts are supported by the following NIH grants: UM1 CA 167552, UM1 CA 186107 and UM1 CA 176726, respectively.

F. K. T., W. W. and E. L. G. designed the study; F. K. T. and W. W. conducted the study and performed statistical analysis; T. T. F., F. B. H., S. A. S.-W., J. E. C., C. S. F. and W. C. W. analysed and interpreted findings and provided critical input; F. K. T. and W. W. wrote the paper; E. L. G. provided study oversight, and had primary responsibility for the final content; all the authors read and approved the final version of the manuscript.

All authors declare that there are no conflicts of interest.

Supplementary material

For supplementary material/s referred to in this article, please visit https://doi.org/10.1017/S0007114516003755

References

1. Yoon Y, Keum N, Zhang X, *et al.* (2015) Hyperinsulinemia, insulin resistance and colorectal adenomas: a meta-analysis. *Metabolism* **64**, 1324–1333.
2. Chen L, Li L, Wang Y, *et al.* (2013) Circulating C-peptide level is a predictive factor for colorectal neoplasia: evidence from the meta-analysis of prospective studies. *Cancer Causes Control* **24**, 1837–1847.
3. Giovannucci E (1995) Insulin and colon cancer. *Cancer Causes Control* **6**, 164–179.
4. Fernández-Veledo S, Nieto-Vazquez I, Vila-Bedmar R, *et al.* (2009) Molecular mechanisms involved in obesity-associated insulin resistance: therapeutical approach. *Arch Physiol Biochem* **115**, 227–239.
5. Kahn B & Flier JS (2000) Obesity and insulin resistance. *J Clin Invest* **106**, 473–481.
6. Reaven GM (1995) Pathophysiology of insulin resistance in human disease. *Physiol Rev* **75**, 473–486.
7. Jensen M, Koh-Banerjee P, Franz M, *et al.* (2006) Whole grains, bran, and germ in relation to homocysteine and markers of glycemic control, lipids, and inflammation. *Am J Clin Nutr* **83**, 275–283.
8. Wu T, Willett WC, Hankinson SE, *et al.* (2005) Caffeinated coffee, decaffeinated coffee, and caffeine in relation to plasma C-peptide levels, a marker of insulin secretion, in U.S. women. *Diabetes Care* **28**, 1390–1396.
9. Hu FB (2002) Dietary pattern analysis: a new direction in nutritional epidemiology. *Curr Opin Lipidol* **13**, 3–9.
10. Jacques PF & Tucker KL (2001) Are dietary patterns useful for understanding the role of diet in chronic disease? *Am J Clin Nutr* **73**, 1–2.
11. Garca-Estévez D, Araújo-Vilar D, Saavedra-González A, *et al.* (2004) Analysis of the relationship between body mass index, insulin resistance, and beta-cell function: a cross-sectional study using the minimal model. *Metabolism* **53**, 1462–1466.
12. Tatsumi Y, Morimoto A, Miyamatsu N, *et al.* (2015) Effect of body mass index on insulin secretion or sensitivity and diabetes. *Am J Prev Med* **48**, 128–135.
13. Dubé JJ, Allison KF, Rousson V, *et al.* (2012) Exercise dose and insulin sensitivity: relevance for diabetes prevention. *Med Sci Sports Exerc* **44**, 793–799.
14. Borghouts L & Keizer HA (2000) Exercise and insulin sensitivity: a review. *Int J Sports Med* **21**, 1–12.

15. Jenkins D, Wolever TM, Taylor RH, *et al.* (1981) Glycemic index of foods: a physiological basis for carbohydrate exchange. *Am J Clin Nutr* **34**, 362–366.
16. Nimptsch K, Brand-Miller JC, Franz M, *et al.* (2011) Dietary insulin index and insulin load in relation to biomarkers of glycemic control, plasma lipids, and inflammation markers. *Am J Clin Nutr* **94**, 182–190.
17. Fung TT, Hu FB, Schulze M, *et al.* (2012) A dietary pattern that is associated with C-peptide and risk of colorectal cancer in women. *Cancer Causes Control* **23**, 959–965.
18. Colditz G & Hankinson SE (2005) The Nurses' Health Study: lifestyle and health among women. *Nat Rev Cancer* **5**, 388–396.
19. Pai J, Pischon T, Ma J, *et al.* (2004) Inflammatory markers and the risk of coronary heart disease in men and women. *N Eng J Med* **351**, 2599–2610.
20. Hankinson SE, Willett WC, Manson JE, *et al.* (1995) Alcohol, height, and adiposity in relation to estrogen and prolactin levels in postmenopausal women. *J Natl Cancer Inst* **87**, 1297–1302.
21. Bonser A & Garcia-Webb P (1984) C-peptide measurement: methods and clinical utility. *Crit Rev Clin Lab Sci* **19**, 297–352.
22. Olson K, Hendricks B & Murdock DK (2012) The triglyceride to HDL ratio and its relationship to insulin resistance in pre- and postpubertal children: observation from the Wausau SCHOOL Project. *Cholesterol* **2012**, 794252.
23. Murguía-Romero M, Jiménez-Flores JR, Sigrist-Flores SC, *et al.* (2013) Plasma triglyceride/HDL-cholesterol ratio, insulin resistance, and cardiometabolic risk in young adults. *J Lipid Res* **54**, 2795–2799.
24. Salazar M, Carbajal HA, Espeche WG, *et al.* (2013) Comparison of the abilities of the plasma triglyceride/high-density lipoprotein cholesterol ratio and the metabolic syndrome to identify insulin resistance. *Diab Vasc Dis Res* **10**, 346–352.
25. McLaughlin T, Abbasi F, Cheal K, *et al.* (2003) Use of metabolic markers to identify overweight individuals who are insulin resistant. *Ann Intern Med* **139**, 802–809.
26. Shai I, Rimm EB, Hankinson SE, *et al.* (2004) Multivariate assessment of lipid parameters as predictors of coronary heart disease among postmenopausal women: potential implications for clinical guidelines. *Circulation* **110**, 2824–2830.
27. Willett W, Stampfer M, Chu NF, *et al.* (2001) Assessment of questionnaire validity for measuring total fat intake using plasma lipid levels as criteria. *Am J Epidemiol* **154**, 1107–1112.
28. Rosner B, Cook N, Portman R, *et al.* (2008) Determination of blood pressure percentiles in normal-weight children: some methodological issues. *Am J Epidemiol* **167**, 653–666.
29. Willett W, Sampson L, Stampfer MJ, *et al.* (1985) Reproducibility and validity of a semiquantitative food frequency questionnaire. *Am J Epidemiol* **122**, 51–65.
30. Feskanich D, Rimm EB, Giovannucci EL, *et al.* (1993) Reproducibility and validity of food intake measurements from a semiquantitative food frequency questionnaire. *J Am Diet Assoc* **93**, 790–796.
31. Rimm E, Giovannucci EL, Stampfer MJ, *et al.* (1992) Reproducibility and validity of an expanded self-administered semiquantitative food frequency questionnaire among male health professionals. *Am J Epidemiol* **135**, 1114–1126.
32. Hu FB, Rimm E, Smith-Warner SA, *et al.* (1999) Reproducibility and validity of dietary patterns assessed with a food frequency questionnaire. *Am J Clin Nutr* **69**, 243–249.
33. Chasan-Taber S, Rimm EB, Stampfer MJ, *et al.* (1996) Reproducibility and validity of a self-administered physical activity questionnaire for male health professionals. *Epidemiology* **7**, 81–86.
34. Wolf AM, Hunter DJ, Colditz GA, *et al.* (1994) Reproducibility and validity of a self-administered physical activity questionnaire. *Int J Epidemiol* **23**, 991–999.
35. Mason C, Foster-Schubert KE, Imayama I, *et al.* (2011) Dietary weight-loss and exercise effects on insulin resistance in postmenopausal women. *Am J Prev Med* **41**, 366–375.
36. Torjesen PA, Birkeland KI, Anderssen SA, *et al.* (1997) Lifestyle changes may reverse development of the insulin resistance syndrome. The Oslo Diet and Exercise Study: a randomized trial. *Diabetes Care* **20**, 26–31.
37. Berger B, Stenström G & Sundkvist G (2000) Random C-peptide in the classification of diabetes. *Scand J Clin Lab Invest* **60**, 687–693.
38. Jones AG & Hattersley AT (2013) The clinical utility of C-peptide measurement in the care of patients with diabetes. *Diabet Med* **30**, 803–817.
39. Palamaner SSG, Kumar AA, Kahan S, *et al.* (2013) Triglyceride/HDL ratio as a screening tool for predicting success at reducing anti-diabetic medications following weight loss. *PLOS ONE* **8**, e69285.
40. Quan H & Zhang J (2003) Estimate of standard deviation for a log-transformed variable using arithmetic means and standard deviations. *Statistics Med* **22**, 2723–2736.
41. Bobe G, Murphy G, Rogers CJ, *et al.* (2010) Serum adiponectin, leptin, C-peptide, homocysteine, and colorectal adenoma recurrence in the Polyp Prevention Trial. *Cancer Epidemiol Biomarkers Prev* **19**, 1441–1452.
42. Dhingra R, Sullivan L, Jacques PF, *et al.* (2007) Soft drink consumption and risk of developing cardiometabolic risk factors and the metabolic syndrome in middle-aged adults in the community. *Circulation* **116**, 480–488.
43. Fung TT, Schulze MB, Hu FB, *et al.* (2012) A dietary pattern derived to correlate with estrogens and risk of postmenopausal breast cancer. *Breast Cancer Res Treat* **132**, 1157–1162.
44. Harris HR, Bergkvist L & Wolk A (2015) An estrogen-associated dietary pattern and breast cancer risk in the Swedish Mammography Cohort. *Int J Cancer* **137**, 2149–2154.
45. Shivappa N, Bosetti C, Zucchetto A, *et al.* (2015) Dietary inflammatory index and risk of pancreatic cancer in an Italian case–control study. *Br J Nutr* **113**, 292–298.
46. Tabung FK, Steck SE, Ma Y, *et al.* (2015) The association between dietary inflammatory index and risk of colorectal cancer among postmenopausal women: results from the Women's Health Initiative. *Cancer Causes Control* **26**, 399–408.
47. Tabung FK, Steck SE, Zhang J, *et al.* (2015) Construct validation of the dietary inflammatory index among postmenopausal women. *Ann Epidemiol* **25**, 398–405.
48. Last AR & Wilson SA (2006) Low-carbohydrate diets. *Am Fam Physician* **73**, 1942–1948.
49. Perrier F, Giorgis-Allemand Li, Slama R, *et al.* (2016) Within-subject pooling of biological samples to reduce exposure misclassification in biomarker-based studies. *Epidemiology* **27**, 378–388.

British Journal of Nutrition