Published by Oxford University Press on behalf of the International Epidemiological Association
International Journal of Epidemiology 2008;37:978–987
doi:10.1093/ije/dyn111

NUTRITION

Associations between dietary methods and biomarkers, and between fruits and vegetables and risk of ischaemic heart disease, in the EPIC Norfolk Cohort Study

Sheila Bingham,[1,2,3]* Robert Luben,[3] Ailsa Welch,[3] Yen Ling Low,[1] Kay Tee Khaw,[3] Nick Wareham[3,4] and Nick Day[3]

Accepted 12 May 2008

Background Methods for assessing diet are prone to measurement error, which may be substantial in large cohort investigations. Biomarkers can be used as objective measures with which to compare estimates of nutritional exposure using different methods

Methods Cross sectional comparisons in 12 474 men and women of regression between biomarkers for vitamin C, sodium, potassium, fibre, carbohydrate, fat and phytoestrogens with intakes derived from food diaries and food frequency questionnaires (FFQ), and odds ratios for risk of ischaemic heart disease (IHD) by dietary and plasma vitamin C.

Results There were strong ($P < 0.001$) associations between biomarkers and intakes as assessed by food diary. Coefficients were markedly attenuated for data obtained from the FFQ, especially so for vitamin C, potassium and phytoestrogens (Z $P < 0.05$). Risk of IHD was associated with plasma vitamin C ($P < 0.001$) and intake of vitamin C and fruit and vegetables assessed by food diary (P quintile trends <0.001, 0.001) but not by the FFQ (P quintile trends 0.923, 0.186).

Conclusions Nutritional data that reflect the findings from biomarkers reduce measurement error and will thus improve statistical power in studies of gene nutrient interactions in cohort studies.

Keywords Diet, food frequency questionnaires, food diaries, biomarkers, EPIC, ischaemic heart disease

Introduction

It is accepted that very large cohort studies are needed for investigating the complex interaction between genetic and environmental factors in common chronic diseases.[1–4]. Nevertheless, although there has been rapid progress in identifying genetic variations that contribute to these diseases, precise and quantitative techniques for the measurement of environmental exposures have not progressed so rapidly.[5] The majority of data relating diet and genetic factors to individual risk of disease in large cohorts relies on a simple method, the food frequency questionnaire (FFQ).

[1] Medical Research Council, Dunn Human Nutrition Unit, Welcome Trust/MRC Building, Cambridge, UK.
[2] MRC Centre for Nutritional Epidemiology in Cancer Prevention and Survival, Institute of Public Health, University of Cambridge, Cambridge, UK.
[3] EPIC Norfolk UK, Institute of Public Health, University of Cambridge, Cambridge, UK.
[4] MRC Epidemiology Unit, Institute of Metabolic Science, Cambridge, UK.
* Corresponding author. MRC Centre for Nutritional Epidemiology, University of Cambridge, Wort's Causeway, Cambridge CBI 8RN. E-mail: sab@mrc-dunn.cam.ac.uk, sheila.bingham@srl.cam.ac.uk

However, misclassification from FFQ leads to attenuation of associations between diet and disease that cannot be corrected even in very large studies.[6–11] This concern has lately been described as a 'crisis' with hundreds of millions of dollars invested in epidemiological studies using the FFQ.[10] The importance of measuring exposure accurately is emphasized by the fact that the subtle effects of gene variants may only become detectable if the presence of certain environmental exposures is also measured.[12,13]

We have previously used biomarkers to assess the accuracy of different dietary measurement techniques used in our cohort study, the Norfolk arm of EPIC (European Prospective Investigation of Cancer).[14] As a result of this validation study with biomarkers we used three different methods to assess diet, a 24 h recall, a FFQ and a detailed food diary kept over 7 days.[15] Increased breast cancer risk was strongly associated with saturated fat intake assessed using the detailed 7 day food diary but not a FFQ.[16] In a larger US-based study, relative risks for breast cancer for women for fat intake assessed by 4 day food records were significant (P for trend 0.02) but not for fat intake when assessed by a FFQ (P for trend 0.24).[17] Measurement error in the assessment of diet is an acknowledged contributor to the controversy surrounding the role of fat in breast cancer risk.[18]

Here, we present data on 12 474 individuals from the EPIC Norfolk cohort who completed both a FFQ and food diary when recruited in 1992–94. We compare the relationship between intake from diet assessed by these two methods with nutritional biomarkers collected in the same individuals. We also show the relationship between dietary vitamin C and fruit and vegetable intake from these two methods compared with plasma vitamin C in risk of ischaemic heart disease (IHD). The results have implications for large investments of public funds in cohort studies of interactions between genetic and nutritional factors within a number of populations such as the UK Biobank, the USA and in Asia.[1–4]

Methods

EPIC Norfolk is a cohort of men and women recruited at age 45–75 years. In 1993, a total of 35 Norfolk medical practices agreed to participate and invited 77 630 individuals to take part in the study. A Health and Lifestyle questionnaire was sent to respondents that was returned with completed informed consent forms by 30 455 participants. The questionnaire included questions on smoking, alcohol consumption, socio-economic status, social class, occupational history, use of medication, dietary changes, history of disease, short family history of main disease end points and reproductive history for women.[14] Exercise was measured by the means of a simple physical activity index, previously validated.[19] Of the participants, 25 639 then agreed to a health examination at which a blood sample, spot urine sample and data on height, weight, respiratory function, anthropometry and blood pressure were collected by trained nurses. All 30 445 participants have been followed up for their health status by flagging with the National Health Service Central Register for death and cancer incidence, the local East Anglian Cancer Registry and the district data base of all in-patient hospital activity. Incident cases of IHD were identified from hospital discharge data or were those who had an underlying cause of IHD (ICD I20-I25) on their death certificate. Those who had been told by their doctor at baseline that they had had a heart attack or stroke were excluded. As part of follow up all participants were invited back for a second health check at the beginning of 1998 at which the protocol was repeated. A total 15 783 individuals attended, of whom 15 025 had attended the first health check. Permission for the study was obtained from The Norfolk and Norwich Hospital Ethics Committee. Consent was provided by the participants for the use of their medical records, to attend a health check and for blood samples to be used at a later date.

Dietary data were obtained using FFQ and 7 day food diaries.[15] The 130-item FFQ was sent with the appointment for the medical examination and were brought to the clinic (usually within about 2 weeks) where they were checked for completeness by nurses at the medical examination. The completed questionnaires were designed to estimate habitual intake over the previous year. Nutrients were computed using an in-house programme, the CAFE (Compositional Analyses from Frequency Estimates) programme.[20] At the medical examination, nurses asked the participants to complete a 7 day food diary comprising an A5, 45 page colour booklet containing food portion photographs and detailed instructions in which the description, preparation and amounts of foods eaten at main meals, snacks and between meals over a week could be recorded.[15,21] Nutrients were calculated using a custom-designed dietary assessment software program, DINER (Data Into Nutrients for Epidemiological Research) program.[21]

Non-fasting serum total cholesterol, high density lipoprotein cholesterol and triglyceride levels were measured with an RA 1000 Technicon analyzer (Bayer Diagnostics, Basingstoke, UK). Urine specimens were frozen without preservative at −20°C. Between 1998 and 2002, the urine samples were thawed and assayed for creatinine on a Roche Cobas Mira Plus analyzer, and sodium and potassium concentrations by flame photometry (IL 943; Instrumentation Lab, Warrington, UK). For vitamin C, 0.25 ml plasma was collected in citrate tubes and stabilized with 0.5 ml 10% metaphosphoric acid prepared fresh weekly. Samples were stored at −70°C until analysis (within 1 week of collection at Addenbrookes Hospital, Cambridge, UK). The plasma ascorbic acid concentration was estimated using a flourometric assay, as

previously described.[22,23] Phytoestrogens were measured by GCMS in urine and LCMS in plasma.[24,25]

Linear regression data between dietary methods and available biomarkers were calculated on 12 474 individuals who completed both a 7 day diary and a FFQ at the first health check. Data on phytoestrogens in plasma and urine was available for 596 individuals.[24,25] Data on diet from the first health check and blood lipids and vitamin C from the second health check were available for 7370 individuals who completed the second health check. Results are shown as mean and 95% CI. Regression coefficients and logistic regression odds ratios were calculated using Stata version 8.2. The results for the trend across quintiles and top vs lower quintile are shown for dietary factors and plasma LDL, HDL, triglycerides and vitamin C and urinary sodium and potassium (corrected for creatinine) and for plasma and urinary (corrected for creatinine) phytoestrogens daidzein and genistein. All biomarker results (except for phytoestrogens that were adjusted only for urinary creatinine) were adjusted for sex, age, smoking, exercise, weight, systolic blood pressure, use of antihypertensive and cholesterol lowering drugs and intake of energy and the energy yielding nutrients alcohol, fat and carbohydrate, except for the plasma LDL and percentage saturated fat energy regressions that were adjusted for carbohydrate only and the plasma HDL and dietary carbohydrate that were adjusted for total fat only. Weight and activity were used as additional adjusters for self reported energy intake, which is poorly measured by dietary methods, particularly FFQ.[26] Hazard ratios (HR) for IHD were adjusted for sex, age, smoking, exercise, weight, systolic blood pressure and intake of energy and the energy-yielding nutrients alcohol, fat and carbohydrate, and for plasma cholesterol. Subjects reporting use of antihypertensive and cholesterol lowering drugs and self reported change in diet over the past year and reporting diabetic, low salt and low fat diet at baseline were excluded.

Results

Tables 1 and 2 show the linear relationships between the intakes of nutrients assessed by both methods and the related biomarker. Trends across quintiles and differences between the top and bottom quintiles of dietary intake data were all $P < 0.0001$ for the data derived from the diary. However, the coefficients were markedly attenuated for data obtained from the FFQ when compared with all biomarkers, especially so for plasma vitamin C and dietary vitamin C, for urinary potassium and dietary potassium when assessed by FFQ and for urinary and serum diadzein and genistein (Z $P < 0.05$), Tables 1 and 2. Dietary and plasma vitamin C associations were similar without adjustment for carbohydrate and fat (coefficient trend diary 5.62 (CI 5.4–5.9) $P < 0.001$, trend FFQ 4.2 (CI 3.9–4.4) $P < 0.001$), Z $P < 0.05$. There was no association between HDL and total fat [for example coefficient Q5 vs Q1 diary 0.014 (CI −0.02 to 0.05) $P = 0.435$], nor between triglycerides and total fat [for example coefficient diary Q5 vs Q1 −0.03 (CI −0.11 to 0.05) $P = 0.458$]. The positive trends between fibre and plasma HDL were weaker than the inverse trends for carbohydrate, for example coefficient Q5 vs Q1 diary 0.027 (CI 0.01–0.05) $P = 0.019$. Alcohol was included throughout the analyses and had significant associations with blood lipids, especially HDL [for example Q1 vs Q5 diary 0.178 (CI 0.16–0.20) $P = <0.001$]. However, protein had no significant effect on blood lipids (for example for HDL Q1 vs Q5 diary −0.007 (CI −0.03 to 0.02) $P = 0.577$ and inclusion of protein did not appreciably affect the magnitude of the regression coefficients [for example for carbohydrate and HDL Q1 vs Q5 diary −0.175 (CI −0.21 to −0.14) $P < 0.001$]. Results shown in the Tables were, therefore, not adjusted for protein. Results for sex specific analyses were consistent with the combined analyses (data not shown).

Table 3 shows comparisons between diet at the first health check and plasma lipids and vitamin C from the second health check 4 years later in the 7370 individuals for whom data was available. Associations between blood lipids and saturated fat, carbohydrate and dietary fibre measured by the diary remained strong, but associations measured by the FFQ were attenuated, as before, and especially so for plasma HDL and carbohydrate (Z $P < 0.05$). Mean plasma vitamin C had increased from 52 (SEM 0.19) at the first health check to 62 (SEM 0.31) mmol/l at the second health check and the relationship between diet and plasma vitamin C assessed by the first health check diary was lessened and the difference between methods was no longer apparent (Z $P > 0.05$).

Table 4 shows quintile means of plasma vitamin C and fruits and vegetables, together with HR in individuals who were free of IHD at the time of recruitment but who later developed IHD. As previously reported, there was an inverse relationship ($P < 0.001$) between first health check plasma vitamin C as a marker for fruit and vegetable intake and risk of IHD.[23]

When vitamin C intake and fruit and vegetable intake were assessed using the food diary, an inverse relationship was observed ($P < 0.001$). However, the relationships between vitamin C and fruit and vegetables and IHD risk were not apparent when assessed by the FFQ. This is despite the almost 2-fold greater apparent intake of fruits and vegetables when assessed by the FFQ compared with the food diary (Table 4). The Figure 1 shows quintile IHD HR estimates for fruits and vegetables and plasma vitamin C.

Discussion

Validation of dietary methods should ideally be carried out using quantitative recovery biomarkers

Table 1 Linear regression data on 12 474 men and women who completed both a 7 day diary and a FFQ at baseline in 1993–1997 in EPIC-Norfolk

	Diary $n = 12\,474$						**FFQ $n = 12\,474$**					
	Trends			Q1 vs Q5			Trends			Q1 vs Q5		
Item	Coefficient	CI	*P* trend	Coefficient	CI	*P* Q1 vs Q5	Coefficient	CI	*P* trend	Coefficient	CI	*P* Q1 vs Q5
Plasma LDL and saturated fat percentage energy	0.028	0.01 to 0.04	<0.001	0.124	0.06 to 0.19	<0.001	0.019	0.01–0.03	0.011	0.065	0.01 to 0.13	0.048
Plasma HDL and dietary carbohydrate	−0.044	−0.05 to −0.04	<0.001	−0.174	−0.21 to −0.14	<0.001	−0.038	−0.05 to −0.03	<0.001	−0.164	−0.21 to −0.12	<0.001
Plasma triglycerides and dietary fibre	−0.031	−0.04 to −0.02	<0.001	−0.141	−0.19 to −0.09	<0.001	−0.025	−0.03 to −0.01	<0.001	−0.110	−0.17 to −0.05	<0.001
Urinary sodium and dietary sodium (mg)	3.996	3.40 to 4.58	<0.001	16.97*	14.4 to 19.5	<0.001	3.089	2.34 to 3.83	<0.001	11.93*	8.74 to 15.11	<0.001
Urinary potassium and dietary potassium (mg)	2.967*	2.61 to 3.32	<0.001	12.82*	11.3 to 14.4	<0.001	1.940*	1.57 to 2.32	<0.001	8.26*	6.66 to 9.86	<0.001
Plasma vitamin C and dietary vitamin C	5.564*	5.33 to 5.79	<0.001	22.97*	21.9 to 24.0	<0.001	3.900*	3.64 to 4.15	<0.001	16.64*	15.5 to 17.8	<0.001

Individuals were categorized into quintiles of the distribution in sex specific nutrient intake by either method. This table shows linear regression results for the trend across quintiles and top vs lower quintile and plasma LDL, HDL, triglycerides and vitamin C and urinary sodium and potassium. All results were adjusted for sex, age, smoking, exercise, weight, systolic blood pressure, use of antihypertensive and cholesterol lowering drugs and intake of energy and the energy yielding nutrients alcohol, fat and carbohydrate, except for the plasma LDL and percentage saturated fat energy regressions, which were adjusted for carbohydrate only and the plasma HDL and dietary carbohydrate which were adjusted for total fat only. Blood samples were collected in a non-fasting state and urines were casual spot samples corrected for creatinine.
*Z $P < 0.05$ β coefficients FFQ vs diary.

Table 2 Linear regression data on 596 men and women who completed both a 7 day diary and a FFQ at baseline in 1993–1997 in EPIC-Norfolk

	Diary						**FFQ**					
	$n = 596$						$n = 596$					
Urinary daidzein and dietary daidzein	0.071*	0.043–0.099	<0.001	0.284	0.158–0.410	<0.001	0.024*	−0.01 to 0.05	0.109	0.112	−0.018 to 0.241	0.090
Urinary genistein and dietary genistein	0.060*	0.035–0.085	<0.001	0.225	0.113–0.336	<0.001	0.016*	−0.01 to 0.04	0.220	0.086	−0.029 to 0.201	0.144
Serum daidzein and dietary daidzein	0.070*	0.039–0.100	<0.001	0.300	0.163–0.438	<0.001	0.021*	−0.01 to 0.05	0.202	0.115	−0.027 to 0.258	0.113
Serum genistein and dietary genistein	0.053*	0.030–0.076	<0.001	0.190	0.086–0.294	<0.001	0.015*	−0.01 to 0.04	0.207	0.080	−0.028 to 0.188	0.146

Individuals were categorized into quintiles of the distribution in sex-specific phytoestrogen intake by either method. This table shows linear regression results for the trend across quintiles and top vs lower quintile and plasma or urinary phytoestrogens. Blood samples were collected in a non-fasting state and urines were casual spot samples corrected for creatinine.
*Z $P < 0.05$ β coefficients FFQ vs diary.

Table 3 Linear regression data on 7401 men and women who completed both a 7 day diary and a FFQ at baseline in 1993–1997 in EPIC-Norfolk and attended a second health check 1997–2000

	Diary $n = 7401$						**FFQ $n = 7401$**					
	Trends			Q1 vs Q5			Trends			Q1 vs Q5		
Item	Coefficient	CI	*P*	Coefficient	CI	*P*	Coefficient	CI	*P*	Coefficient	CI	*P*
Plasma LDL and saturated fat percentage energy	0.032	0.01 to 0.05	0.001	0.135	0.05 to 0.22	0.002	0.009	−0.01 to 0.03	0.360	0.018	−0.07 to 0.10	0.675
Plasma HDL and dietary carbohydrate	−0.043*	−0.06 to −0.03	<0.001	−0.176	−0.23 to −0.12	<0.001	−0.026*	−0.04 to −0.01	<0.001	−0.107	−0.17 to −0.05	0.001
Plasma triglycerides and dietary fibre	−0.027	−0.05 to −0.01	0.004	− 0.121	−0.20 to −0.04	0.004	−0.020	−0.04 to −0.00	0.044	−0.111	−0.20 to −0.02	0.013
Plasma vitamin C and dietary vitamin C	3.39	2.97 to 3.81	<0.001	13.81	11.9 to 15.7	<0.001	3.16	2.73 to 3.59	<0.001	13.28	11.4 to 15.2	<0.001

Individuals were categorized into quintiles of the distribution in sex specific nutrient intake by either method. This table shows linear regression results for the trend across quintiles and top vs lower quintile and plasma LDL, HDL, triglycerides and vitamin C from bloods taken at the second health check. All results were adjusted as in Table 1. Blood samples were collected in a non-fasting state.
*Z $P < 0.05$ β coefficients FFQ vs diary.

Table 4 Quintile means and HR per quintile increase in variable using Cox regression for vitamin C in 678 cases of IHD (mortality and incident hospital admissions) in 11 134 individuals who completed both a 7 day diary and a FFQ

	Q1	Q2	Q3	Q4	Q5
Quintile means					
Plasma vitamin C mmol/l	28.9	46.3	55.1	63.1	79.9
Fruit and vegetables (gm/day)					
Food diary	79	160	226	306	486
FFQ	200	324	422	536	824

	Trends			Q1 vs Q5		
Item	HR	CI	*P* trend	HR	CI	*P* Q1 vs Q5
Plasma vitamin C and IHD	0.900	0.85–0.95	<0.001	0.629	0.48–0.82	0.001
Diary						
Diet vitamin C and IHD	0.910	0.86–0.96	0.001	0.616	0.47–0.81	<0.001
Diet fruit and vegetables and IHD	0.904	0.85–0.96	0.001	0.632	0.48–0.82	0.001
FFQ						
Diet vitamin C and IHD	0.997	0.94–1.05	0.923	0.860	0.66–1.11	0.225
Diet fruit and vegetables and IHD	0.961	0.90–1.01	0.186	0.901	0.70–1.67	0.431

All HR were adjusted for sex, age, smoking, exercise, weight, systolic blood pressure and intake of energy and the energy yielding nutrients alcohol, fat and carbohydrate, and for plasma cholesterol. Subjects reporting use of antihypertensive and cholesterol lowering drugs and self reported change in diet over the past year and reporting diabetic, low salt and low fat diet at baseline were excluded.

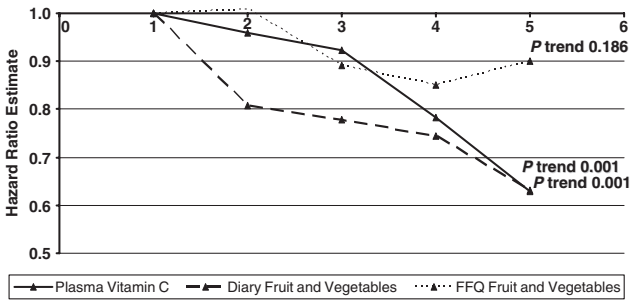

Figure 1 Plasma vitamin C and fruit and vegetable intake from different methods and risk of developing IHD in 678 cases in EPIC Norfolk Cohort of 11 134 free of CHD at baseline. HR per quintile are shown

such as doubly labelled water or 24 h urine collections.[27] We used this approach to inform our early decision to obtain information using three different methods in the EPIC Norfolk cohort.[28] The availability and expense of quantitative recovery biomarkers made their use not possible on this, the largest population comparison sample of dietary methods incorporating biomarkers ever studied. However, plasma vitamin C, plasma and spot urinary phytoestrogens, and spot urinary sodium and potassium are well established comparison biomarkers of dietary exposure, because they are correlated with dietary intake on an individual basis even though quantitative recoveries are not obtained.[27]

LDL has not previously been suggested as a biomarker of fat intake since it is affected by a number of other factors besides diet. Nonetheless, carefully controlled metabolic and cross sectional studies have established that percentage energy from saturated fatty acids is one of the major determinants of serum LDL and total cholesterol.[29,30] We have shown elsewhere in EPIC Norfolk that associations with plasma LDL and total saturated fat and percentage energy from total saturated fat were similar and we have used the convention of percentage energy from saturated fat here.[31] Investigations of the effect of fibre have previously concentrated on its effect on plasma cholesterol rather than on serum triglyceride levels[32] and there are conflicting results from intervention studies, with some studies showing no effects and others inverse effects of fibre on serum triglycerides.[33–35] Carbohydrate has not previously been suggested as a marker of triglycerides, but carbohydrate typically increases triglyceride concentrations more than does fat, which generally does not raise serum triglyceride levels.[36] Effects of carbohydrate in this study were, however, weaker than those of fibre. Cross sectional inverse associations between fibre intake and serum triglycerides have been shown in another prospective study, the Framingham study of women.[37] We measured serum lipids in the non-fasting state. However, blood sampling would have had no major effect on our analysis as indicated in a meta-analysis that showed no differences in triacylglycerol-associated IHD risk between non-fasting and fasting participants.[38] HDL has previously been suggested as a marker of fat intake.[39] In the

present study, we found no association between HDL and total fat, but we did find a strong inverse association with carbohydrate. Fat is inversely associated with carbohydrate intake and it is well established in the literature that isoenergetic substitution of saturated, monounsaturated and polyunsaturated fatty acids (but not *trans* fatty acids) with carbohydrate lowers HDL levels.[36,40–42] Lower HDL concentrations are also found in both individuals and populations that habitually consume low fat, high carbohydrate diets.[36,40,42]

Using these comparison markers for exposure to fat, carbohydrate, fibre, vitamin C, potassium, sodium and phytoestrogens, we show that there were strong and consistent associations between biomarkers and dietary intake when assessed by the food diary. However, associations with biomarkers were attenuated when intake was assessed by the FFQ (Tables 1 and 2). Table 3 shows that the associations with fat, fibre and carbohydrate biomarkers remained strong for dietary intake measured by the food diary 4 years earlier but were attenuated when dietary intake was measured 4 years earlier by the FFQ, especially so for carbohydrate and plasma HDL. There were no differences in coefficients of associations between the biomarkers and intakes of vitamin C from different methods using second health check data (Table 3). However, plasma vitamin C had increased between the two health checks, likely due to an increase in dietary intake, as national data show that there has been a substantial (30–35%) increase in dietary intake of vitamin C between 1987–2000.[43] A change in intake, rather than loss of power from a reduction in sample size by 40%, thus explains why the relationship between intake from the first health check diary and plasma vitamin C at the second health check was lessened. It is noteworthy that the magnitudes of the *b* coefficients between intake of vitamin C from the FFQ and plasma vitamin C were similar on both occasions, as was the attenuated relation between the intake from the food diary and the changed plasma vitamin C up to 4 years later. This suggests that the association between plasma vitamin C and intake of vitamin C from the FFQ was less specific than the relation between the actual intake measured from the diary and plasma vitamin C at the first health check. Unfortunately no coded second health check diaries are available as yet to confirm this point. Plasma vitamin C is most highly correlated with dietary vitamin C for up to 30 days preceeding blood collection ($r = 0.5$–0.6), within the time the FFQ was generally completed and blood was taken at the first health check.[44] This, taken together with the data in Table 3 suggests there is little evidence that the results shown in the Tables 1 and 2 can be attributed to the fact that the collection of data from the biomarkers was closer in time to the food diary (within 1 week) than that from the FFQ (completed before attending the clinic where biomarkers were collected).

We have previously shown that plasma vitamin C is associated with reduced risk of IHD. Plasma vitamin C is a biomarker of fruits and vegetable consumption and unlikely to be the active protective factor, since supplements of pure vitamin C have had no effect on IHD mortality.[23] It has been suggested that, as socioeconomic factors are also related to plasma vitamin C, these could account for associations between plasma vitamin C and fruits and vegetables with IHD risk.[45] However, as data on all exposures were available on the same individuals, it is unlikely that the apparent lack of effect of fruits and vegetables on IHD risk when assessed by FFQ is accounted for by socioeconomic factors, especially as the FFQ is recommended because it is less demanding for participants to complete.[10] For consistency, we used the same adjusting factors for energy and risk factors when comparing methods with biomarkers, and assessing HR for IHD, throughout.

Table 4 shows that the association between risk of IHD and consumption of fruits and vegetables would not have been evident had a FFQ been used. However, using the food diary, there was a relative risk reduction of 37%, 0.63 (0.48 – 0.82) in the top quintile of fruit and vegetable intake (equivalent to six portions per day) compared with the bottom quintile. This is of a much greater magnitude than that of the relative risk reduction of only 12% for an increment of five servings daily of 0.88 (95% CI 0.81–0.95) for cardiovascular disease in a large US study of 3634 cardiovascular cases in a cohort of 109 635 individuals in which intake was determined using a FFQ.[46] In the present study, and in the US study, fruits and vegetables assessed by the FFQ were remarkably high. It is difficult to compare results between populations quantitatively but findings from the FFQs might suggest that the average population health target of five portions or fruit and vegetables (400 g/day) had been surpassed. However, FFQ are well known to systematically overestimate fruits and vegetable intakes[15] and Table 4 shows that less than three portions per day were consumed on average in the middle quintile when assessed by the food diary, an amount similar to that derived from UK national food survey average data.[43]

We have been able to assess associations between biomarkers and dietary intake using different methods using a very limited range of available biomarkers, and our findings may differ for other nutrients and foods. For example in a smaller comparison of 4949 individuals in EPIC Norfolk, the correlations between n-3 plasma fatty acids and fatty fish intake assessed by food diary and FFQ were the same,[47] whereas in our earlier study using recovery biomarkers, intake of protein and potassium as estimated by the food diary was more closely related to 24 h urine biomarkers for protein and potassium ($r = 0.60$–0.70) than when estimated by FFQ ($r = 0.27$–0.50).[48] In the OPEN study, correlations between the biomarkers

doubly labelled water and 24 h urine nitrogen with energy intake and protein intake assessed by two 24 h recalls were higher ($r = 0.24$–0.41) than when assessed by FFQ ($r = 0.10$–0.33).[49] These and other validation studies investigating the relationship between biomarkers and different methods are, however, generally small, involving <500 subjects,[6,50,51] so that investigation of the predictive value of different methods in relation to disease end points is not possible. The present set of data are, however, sufficiently large to compare plasma vitamin C, a biomarker of intake of fruits and vegetables, and intake from different methods of dietary assessment in relation to risk of IHD risk.

Energy adjustment has been recommended to attempt to correct for measurement error, but effects can vary widely depending on the nutrient concerned, with negligible effects in the case of vitamin C, but with greater and unpredictable effects for items with a high error correlation such as fat.[52] In addition, total energy intake is imperfectly measured and we, therefore, adjusted for weight and activity throughout, as these are more closely related to energy expenditure and, in energy balance, energy intake.[26,53]

Data shown here comparing different dietary methods suggest that those methods most closely related to biomarkers are associated with an improved association between diet and chronic disease, in this case fruit and vegetable intake and IHD, probably due to reduced measurement error and thus improved accuracy. Improved accuracy is a cost effective way of avoiding substantial infrastructure requirements for recruitment and follow up of massive population cohorts in which interaction between gene variants and dietary or other environmental variable exposure is to be assessed.[53] Records, recalls or other ways of documenting real time intake are more expensive to analyse, but written records can be stored for later analysis in nested case control studies, as can biological markers. Self administered CD and web-based methods are under development. The advantage of heterogeneity and calibration from multi-cohort studies in food habits can also be used to improve accuracy, at least for some items, as in the Europe wide EPIC study.[54]

Acknowledgements

EPIC Norfolk is funded by the Medical Research Council, Cancer Research UK, Food Standards Agency, World Cancer Research Fund, British Heart Foundation, Yen Ling Low by the Agency for Science, Technology and Research (Centros, Singapore).

Conflict of interest: None declared.

KEY MESSAGES

- In assessing diet in cohort investigations of gene nutrient interactions in chronic disease risk, a simple method, a FFQ, is usually used.
- In this study of 12 474 men and women, we used a second method, a food diary, to assess dietary intake.
- Cross sectional associations between dietary intake derived from the food diary and nutritional biomarkers in blood and urine were stronger than those derived from the FFQ.
- Risk of IHD was associated with plasma vitamin C ($P < 0.001$) and intake of vitamin C and fruit and vegetables assessed by food diary (P quintile trends <0.001, 0.001) but not by the FFQ (P quintile trends 0.923, 0.186).

References

[1] Willett WC, Collins FS, Manolio TA. Merging and emerging cohorts. *Nature* 2007;**445:**257–58.

[2] Cyranoski D, Williams R. Health study sets sights on a million people. *Nature* 2005;**434:**812.

[3] Francis S, Collins FS. The case for a US prospective cohort study of genes and environment. *Nature* 2004;**429:**475.

[4] Barbour V. UK Biobank: a project in search of a protocol? *Lancet* 2003;**361:**1734–38.

[5] Schartz D, Collins F. Environmental biology and human disease. *Science* 2007;**316:**695–96.

[6] Day NE, McKeown N, Wong MY, Welch A, Bingham S. Epidemiological assessment of diet: a comparison of a 7 day diary with a food frequency questionnaire. *Int J Epidemiol* 2001;**30:**309–31.

[7] Kipnis V, Subar AF, Midthune D *et al*. Structure of dietary measurement error: results of the OPEN biomarker study. *Am J Epidemiol* 2003;**158:**14–21.

[8] Prentice RL. Dietary assessment and the reliability of nutritional epidemiology reports. *Lancet* 2003;**362:** 182–83.

[9] Schatzkin A, Kipnis V. Could exposure assessment problems give us wrong answers to nutrition and cancer questions? *J Natl Cancer Inst* 2004;**96:**1564–65.

[10] Kristal AR, Peters U, Potter JD. Is it time to abandon the food frequency questionnaire? *Cancer Epidemiol Biomarkers Prev* 2005;**14:**2826–28.

[11] Freedman LS, Schatzkin A, Thiebaut ACM *et al*. Abandon neither the food frequency questionnaire nor the dietary fat-breast cancer hypothesis. *Cancer Epidemiol Biomarkers Prev* 2007;**16:**1321–22.

[12] Clayton D, McKeigue PM. Epidemiological methods for studying genes and environmental factors in complex diseases. *Lancet* 2001;**358:**1356–60.

[13] Low Y-L, Dunning AM, Dowsett M *et al*. Implications of gene-environment interaction in studies of gene variants in breast cancer: an example of dietary isoflavones and the D356N polymorphism in the sex hormone-binding globulin gene. *Cancer Res* 2006;**66:**8980–83.

[14] Day NE, Oakes S, Luben R *et al*. EPIC in Norfolk: study design and characteristics of the cohort. *Br J Cancer* 1999;**80 (Suppl 1):**95–103.

[15] Bingham S, Welch A, McTaggart A *et al*. Nutritional methods in the European Prospective Investigation of Cancer in Norfolk. *Public Health Nutr* 2001;**4:**847–58.

[16] Bingham SA, Luben R, Welch A, Wareham N, Khaw KT, Day NE. Fat and breast cancer: are imprecise methods obscuring a relationship? Report from the EPIC Norfolk prospective cohort study. *Lancet* 2003;**362:**212–14.

[17] Freedman LS, Potischman N, Kipnis V *et al*. A comparison of two dietary instruments for evaluating the fat-breast cancer relationship. *Int J Epidemiol* 2006;**35:**1011–21.

[18] Thiebaut ACM, Kipnis V, Schatzkin A, Freedman LS. The role of dietary measurement error in investigating the hypothesized link between dietary fat intake and breast cancer - A story with twists and turns. *Cancer Invest* 2008;**26:**68–73.

[19] Wareham NJ, Jakes RW, Rennie KL, Mitchell J, Hennings S, Day NE. Validity and repeatability of the EPIC-Norfolk Physical Activity Questionnaire. *Int J Epidemiol* 2002;**31:**168–74.

[20] Welch AA, Luben R, Khaw KT, Bingham SA. The CAFE computer program for nutritional analysis of the EPIC-Norfolk Food Frequency Questionnaire and identification of extreme values. *J Hum Nutr Diet* 2005;**18:**99–116.

[21] Welch AA, McTaggart A, Mulligan AA *et al*. DINER (Data Into Nutrients for Epidemiological Research) - a new data-entry program for nutritional analysis in the EPIC-Norfolk cohort and the 7-day diary method. *Public Health Nutr* 2001;**4:**1253–65.

[22] Vuilleumier J, Keck E. Fluorometric assay of vitamin C in biological materials using a centrifugal analyser with fluorescence attachment. *J Micronutr Anal* 1989;**5:**25–34.

[23] Khaw KT, Bingham S, Welch A *et al*. Relation between plasma ascorbic acid and mortality in men and women in EPIC-Norfolk prospective study. *Lancet* 2001;**357:**657–63.

[24] Low YL, Taylor JI, Grace PB *et al*. Polymorphisms in the CYP19 gene may affect the positive correlations between serum and urine phytoestrogen metabolites and plasma androgen concentrations in men. *J Nutr* 2005;**135:**2680–86.

[25] Grace PB, Taylor JI, Low YL *et al*. Phytoestrogen concentrations in serum and spot urine as biomarkers for dietary phytoestrogen intake and their relation to breast cancer risk in EPIC-Norfolk. *Cancer Epidemiol Biomarkers Prev* 2004;**13:**698–709.

[26] Jakes RW, Day NE, Luben R *et al*. Adjusting for energy intake - what measure to use in nutritional epidemiological studies? *Int J Epidemiol* 2004;**33:**1382–86.

[27] Bingham SA. Biomarkers in nutritional epidemiology. *Public Health Nutr* 2002;**5:**821–28.

[28] Bingham SA, Cassidy A, Cole T *et al*. Validation of weighed records and other methods of dietary assessment using the 24 h urine nitrogen technique and other biological markers. *Br J Nutr* 1995;**73:**531–50.

[29] Hegsted DM, Ausman LM, Johnson JA, Dallal GE. Dietary fat and serum lipids: an evaluation of the experimental data. *Am J Clin Nutr* 1993;**57:**875–83.

[30] Keys A, Anderson JT, Grande F. Prediction of serum-cholesterol responses of man to changes in fats in the diet. *Lancet* 1957;**273:**959–66.

[31] Wu K, Bowman R, Welch A *et al*. Apoliporotein E polymorphisms, dietary fat and fibre and serum lipids: the EPIC Norfolk study. *Eur Heart J* 2007;**28:** 2930–36.

[32] Truswell AS. Dietary fibre and plasma lipids. *Eur J Clin Nutr* 1995;**49 (Suppl 3):**S105–9.

[33] Hunninghake DB, Miller VT, Larosa JC *et al*. Long-term treatment of hypercholesterolemia with dietary fiber. *Am J Med* 1994;**97:**504–8.

[34] Panlasigui LN, Baello OQ, Dimatangal JM, Dumelod BD. Blood cholesterol and lipid-lowering effects of carrageenan on human volunteers. *Asia Pac J Clin Nutr* 2003;**12:**209–14.

[35] Zunft HJF, Luder W, Harde A *et al*. Carob pulp preparation rich in insoluble fibre lowers total and LDL cholesterol in hypercholesterolemic patients. *Eur J Nutr* 2003;**42:**235–42.

[36] Grundy SM, Denke MA. Dietary influences on serum lipids and lipoproteins. *J Lipid Res* 1990;**31:**1149–72.

[37] Sonnenberg LM, Quatromoni PA, Gagnon DR *et al*. Diet and plasma lipids in women. II. Macronutrients and plasma triglycerides, high-density lipoprotein, and the ratio of total to high-density lipoprotein cholesterol in women: the Framingham Nutrition studies. *J Clin Epidemiol* 1996;**49:**665–72.

[38] Sarwar N, Danesh J, Eiriksdottir G *et al*. Triglycerides and the risk of coronary heart disease: 10,158 incident cases among 262,525 participants in 29 Western prospective studies. *Circulation* 2007;**115:**450–58.

[39] Willett W, Stampfer M, Chu NF, Spiegelman D, Holmes M, Rimm E. Assessment of questionnaire validity for measuring total fat intake using plasma lipid levels as criteria. *Am J Epidemiol* 2001;**154:**1107–12.

[40] Mensink RP, Zock PL, Kester ADM, Katan MB. Effects of dietary fatty acids and carbohydrates on the ratio of serum total to HDL cholesterol and on serum lipids and apolipoproteins: a meta-analysis of 60 controlled trials. *Am J Clin Nutr* 2003;**77:**1146–55.

[41] Katan MB, Zock PL, Mensink RP. Effects of fats and fatty-acids on blood-lipids in humans - an overview. *Am J Clin Nutr* 1994;**60:**S1017–22.

[42] Katan MB. Effect of low-fat diets on plasma high-density lipoprotein concentrations. *Am J Clin Nutr* 1998;**67:** 573S–76S.

[43] Henderson L, Irving K, Gregory J *et al. The National Diet and Nutrition Survey: Adults Aged 19 to 64 Years*. Vol. 3. London: HMSO, 2003.

[44] Bates CJ, Thurnham DI, Bingham SA, Margetts BM, Nelson M. Biochemical markers of nutrient intake. In: Margetts BM, Nelson M (eds). *Design Concepts in Nutritional Epidemiology*. 2nd edn. Oxford: Oxford University Press, 1997, pp. 170–240.

[45] Lawlor DA, Davey Smith G, Kundu D, Bruckdorfer KR, Ebrahim S. Those confounded vitamins: what can we

learn from the differences between observational versus randomised trial evidence? *Lancet* 2004;**363:**1724–27.

[46] Hung HC, Joshipura KJ, Jiang R *et al*. Fruit and vegetable intake and risk of major chronic disease. *J Natl Cancer Inst* 2004;**96:**1577–84.

[47] Welch AA, Bingham SA, Ive J *et al*. Dietary Fish Intake and plasma phospholipid n-3 polyunsaturated fatty acid concentrations in EPIC Norfolk. *Am J Clin Nutr* 2006;**84:** 1330–39.

[48] Bingham S, Gill C, Welch A *et al*. Validation of dietary assessment methods in the UK arm of EPIC. *Int J Epidemiol* 1997;**26:**S137–51.

[49] Subar AF, Kipnis K, Troiano RP *et al*. Using intake biomarkers to evaluate the extent of dietary misreporting in a large sample of adults: The Observing Protein and Energy Nutrition (OPEN) study. *Am J Epidemiol* 2003; **158:**1–13.

[50] Kabagambe EK, Baylin A, Allan DA, Siles X, Spiegelman D, Campos H. Application of the method of triads to evaluate the performance of food frequency questionnaires and biomarkers as indicators of long-term dietary intake. *Am J Epidemiol* 2001;**154:** 1126–35.

[51] Shai I, Rosner BA, Shahar DR *et al*. Dietary evaluation and attenuation of relative risk: multiple comparisons between blood and urinary biomarkers, food frequency, and 24-hour recall questionnaires: the DEARR study. *J Nutr* 2005;**135:**573–79.

[52] Day NE, Wong MY, Bingham SA *et al*. Correlated measurement error- implications for nutritional epidemiology. *Int J Epidemiol* 2004;**33:**1–9.

[53] Wong MY, Day NE, Luan JA, Wareham NJ. Estimation of magnitude in gene-environment interactions in the presence of measurement error. *Stat Med* 2004;**23:** 987–98.

[54] Bingham S, Riboli E. Is diet important in cancer pathogenesis? The European Prospective Investigation of Cancer (EPIC). *Nat Rev Cancer* 2004;**4:**206–15.

Published by Oxford University Press on behalf of the International Epidemiological Association

International Journal of Epidemiology 2008;**37**:987–989
doi:10.1093/ije/dyn177

Commentary: Flawed study designs are not salvaged by large samples

Walter C Willett

Accepted 29 July 2008

In this issue, Bingham *et al*.[1] present analyses comparing nutrient intakes estimated by a 1-week diet record (DR) or food frequency questionnaire (FFQ) to a series of biomarkers and incidence of coronary heart disease (CHD). Such triangulation analyses can provide valuable evidence on the relative validity of different dietary assessment methods and are thus of considerable interest. Unfortunately, the results provided by Bingham *et al*. are difficult to interpret due to the design of the study and their choice of statistical models.

In the design of a study in which multiple dietary assessment methods are compared with biomarkers, the temporal relationships are critical. Underlying this is the concept that, for epidemiologic applications, the dietary methods are meant to represent intake over an extended period, usually many months or years, not just the few days or weeks around the time the data are collected. Thus, if the biomarker is sensitive to recent intake (true for most urinary and plasma measurements, but less true for biomarkers using adipose or nail tissue) and is collected close in time to a measure of short-term intake, such as a DR or several 24-h recalls, the correlations will tend to be exaggerated. This is a form of correlated error that can result from fluctuations in diet due to seasonality or random variation that will be reflected in both measurements (see Chapter 6).[2] These issues are different for a FFQ, which typically asks about diet over the past year, and should therefore reflect dietary intake over a much longer period, not just recent diet. In the cross-sectional analyses by Bingham *et al*., the FFQ, DR and biomarkers were all collected in close temporal proximity, which would tend to bias the correlations in favour of the DR. Their argument in the discussion that the FFQ was collected within 30 days of the biomarkers misses the critical point. The solution to this bias is to collect the biomarker at some time remote from the collection of the dietary intakes assessments, either before or later. Fortunately, Bingham *et al*. have done this. In an earlier, although smaller, validation study, plasma vitamin C

Department of Nutrition, Harvard School of Public Health, 665 Huntington Avenue, Boston, MA 02115, USA.
E-mail: walter.willett@channing.harvard.edu