Biomarkers, July–August 2006; 11(4): 370–382

informa
healthcare

Reliability of biomarkers of iron status, blood lipids, oxidative stress, vitamin D, C-reactive protein and fructosamine in two Dutch cohorts

W. K. AL-DELAIMY[1,4], E. H. J. M. JANSEN[2], P. H. M. PEETERS[3], J. D. VAN DER LAAN[2], P. A. H. VAN NOORD[3], H. C. BOSHUIZEN[2], Y. T. VAN DER SCHOUW[3], M. JENAB[1], P. FERRARI[1], & H. B. BUENO-DE-MESQUITA[2]

[1]*International Agency for Research on Cancer (IARC-WHO), Lyon, France,* [2]*National Institute for Public Health and the Environment (RIVM), Bilthoven, the Netherlands,* [3]*Julius Center for Health Sciences and Primary Care, University Medical Center, Utrecht, the Netherlands and* [4]*Moores UCSD Cancer Center, University of California, San Diego, CA, USA*

Abstract

Biomarkers are widely used in epidemiology, yet there are few reliability studies to assess the appropriateness of using these biomarkers for the assessment of exposure–disease relationships. The aim of the study was to assess the reliability of 20 biomarkers in serum collected from two Dutch centres (Utrecht and Bilthoven) participating in the European Prospective Investigation into Cancer and Nutrition (EPIC) at two points several years apart. Blood samples were collected from 30 men from Bilthoven and 35 women from Utrecht. Ferritin, total iron, total iron-binding capacity, transferrin saturation, transferrin, C-reactive protein, bilirubin, cholesterol, triglycerides, apo lipoprotein-A, apo lipoprotein-B, high-density lipoproteins, low-density lipoproteins, uric acid, creatinine, reactive oxygen metabolites, the ferric-reducing ability of plasma, protein thiol oxidation, fructosamine, and vitamin D biomarkers in serum were analysed from the blood samples at the two points of time. For all biomarkers, except C-reactive protein, there were no substantial changes in the mean levels over time. Uric acid, ferritin, creatinine, HDL, and apo lipoprotein-B levels consistently showed the highest reliability for men and women (intra-class correlation = 0.69–0.86). Among women, the ferric-reducing ability of plasma, and protein thiol oxidation had poor reliability; and among men iron-related biomarkers (except serum ferritin) had poor reliability. With the exception of a few gender-specific differences, most of the 20 biomarkers performed well and can be considered to have sufficient reliability to be used in future cohort studies.

Keywords: *Reliability, biomarkers, cohort, European Prospective Investigation into Cancer and Nutrition (EPIC)*

(Received 2 February 2006; accepted 11 May 2006)

Correspondence: W. Al Delaimy, Cancer Prevention and Control Program, Rebecca and John Moores UCSD Cancer Center, University of California, San Diego, 3855 Health Sciences Drive, La Jolla, CA 92093-0901, USA. Tel: 1-858-822-6515. Fax: 1-858-822-2399. E-mail: wael@ucsd.edu

ISSN 1354-750X print/ISSN 1366-5804 online
DOI: 10.1080/13547500600799748

Introduction

Assessing the reliability (also known as repeatability as an indicator of consistent results over time) of any exposure measurement methods is important before carrying out large epidemiological investigations. Unreliable measures may bias relative risk estimates toward the null value and reduce the effective sample size and statistical power of a study (Shrout & Fleiss 1979, De Klerk et al. 1989, McKeown-Eyssen & Tibshirani 1994, Willett 1998). Similarly, if biomarkers of dietary intake, internal metabolites, external exposures or predictors of disease outcomes fail to reflect consistently the factor of interest, such biomarkers will be of limited value in epidemiological research and their use will be a waste of stored samples and the resources involved in collecting and analysing such samples.

The issue of biomarker reliability is particularly important for cohort studies using prospectively collected data, where samples (usually blood) are collected at one time point or only from a subpopulation due to cost or logistical constraints. The reliability of a biomarker is dependent on the reliability of the laboratory assay used for its measure, as well as the reliability of the measurement of interest within the same participant over time. Therefore, a highly reliable biomarker must have limited laboratory assay variability and can consistently reflect the factor of interest in a population over time. Biological variability of the factor of interest cannot be measured or controlled but will be reflected by low reliability of that factor when there is high biological variability. Confounding variables and effects that influence biomarker levels should be assessed simultaneously with collection of the biomarker sample. These factors should then be adjusted for during data analyses, or those individuals whose factors have changed over time should be excluded from the study population.

There is limited knowledge about the reliability of some of the important biomarkers of oxidative damage, inflammation, or nutrient intake. The present study assessed the reliability of 20 such biomarkers over time which are of relevance to many chronic diseases and have established laboratory assays that can be carried out with good validity and reliability. Their overall reliability will determine their use in the large cohort of the European Prospective Investigation into Cancer and Nutrition (EPIC). The selected biomarkers are relevant to colorectal cancer (CRC) (Kato et al. 1999, Gackowski et al. 2002, Erlinger et al. 2004, Grant & Garland 2004, Misciagna et al. 2004, Poynter et al. 2005) as well as other diseases. These biomarkers include the following:

1. Iron status: serum ferritin, total serum iron, total iron-binding capacity (TIBC), transferrin saturation, and transferrin (TRSF).
2. Inflammation: C-reactive protein (CRP).
3. Blood lipids: cholesterol, triglycerides, apo lipoprotein-A, apo lipoprotein-B, high-density lipoproteins (HDL), and low-density lipoproteins (LDL) serum levels.
4. Markers of oxidative stress: uric acid, creatinine, bilirubin, reactive oxygen metabolites (ROM), the ferric-reducing ability of plasma (FRAP), and protein thiol oxidation (PSH).
5. Glucose metabolism: fructosamine.
6. 25-Hydroxy vitamin D.

The data for this study were from participants in a sub-sample of the Dutch cohorts of EPIC who donated blood samples and provided questionnaire information at two different time points several years apart. The aim was to assess the reliability of the above biomarkers in this sub-sample of participants.

Materials and methods

The European Prospective Investigation into Cancer and Nutrition (EPIC) is a large European prospective cohort study consisting of 23 centres from ten European countries (Riboli et al. 2002, Bingham & Riboli 2004). The present study is based on data collected from the two Dutch cohorts (Bilthoven and Utrecht) participating in EPIC. Data were collected from participants who provided blood samples at two different time points from each of the Dutch centres to compare biomarkers from their blood samples collected at these two points in time. There were 235 male participants from Bilthoven and 501 female participants from Utrecht who had available blood samples collected at two separate time intervals. Other data collected at these time points were disease history, weight, smoking, alcohol intake, energy intake, and vitamin intake, all of which have an effect on the serum levels of one or more of the biomarkers of interest (Smit et al. 1994, Ocké et al. 1997a Ocké et al. 1997b, Boker et al. 2001). Participants who did not meet the selection criteria were not selected in order to avoid possible confounding on the reliability of the analysed biomarkers (selection criteria summarized in Appendix 1). Local ethics committees approved both cohorts.

For Bilthoven, 487 Doetinchem town participants in the Morgen-EPIC cohort (235 men and 252 women) were recruited between May 1993 and February 1994 (this is the baseline that is referred to as Time 1) (Smit et al. 1994, Riboli & Kaaks 1997, Van Loon et al. 2003). After approximately 5 years (between May 1998 and February 1999) they were subjected to a complete repeat of the baseline examination (which will be referred to as Time 2), including a repeat blood sample for the purpose of intra-individual variance estimates. The repeat blood sample was collected in the same season as those that were drawn at baseline. The study population for the analyses is a sub-sample of the Doetinchem participants in the Morgen-EPIC cohort study. For the purpose of comparability, only men were selected from the Bilthoven cohort, since the other cohort of Utrecht had only women participants.

For Utrecht, women between ages 50 and 69 years, who lived in the city of Utrecht and the nearby vicinity and were scheduled for breast cancer screening from 1993 to 1997 (baseline or Time 1), received a mailed invitation to join the Prospect-EPIC study (De Waard et al. 1984, Fracheboud et al. 1998, Boker et al. 2001) with their invitation for a routine mammography. A second examination and blood collection were performed for approximately 700 women participants after 2–3 years (Time 2). We ended up selecting 30 men from Bilthoven and 35 women from Utrecht for this study based on the selection criteria described in the Appendix.

Blood collection and laboratory assays

As part of the standardized blood collection protocol of all EPIC participants, non-fasting venous blood samples were obtained using 3 × 10-ml Safety-Monovettes tubes (Sarstedt, Tilburg, the Netherlands) from each participant in Bilthoven and Utrecht.

In both centres, filled syringes were kept at 5–10°C, protected from light, and transferred in cool boxes to the respective Cryo-laboratory of the National Institute for Public Health and the Environment (RIVM), Bilthoven. For the Bilthoven participants for this study, the blood samples stored at the RIVM were used. For the Utrecht participants, blood samples transported in liquid nitrogen vapour to the International Agency for Research on Cancer (IARC) and again stored under liquid nitrogen were used. After retrieval, serum samples of the Utrecht participants were transported under dry ice from the IARC to the biochemical laboratory of RIVM in Bilthoven. Further details of sample preparation are described elsewhere (Al-Delaimy et al. 2004).

From 14 to 21 June 2004, the following biomarkers were determined with a clinical chemical autoanalyser (Hitachi 912, Roche Diagnostics, Mannheim, Germany): uric acid, creatinine, bilirubin, triglycerides, total cholesterol, HDL-cholesterol, apo lipoprotein-A and -B, total serum iron, transferrin (TRSF), fructosamine, and high-sensitivity C-reactive protein (CRP). LDL-cholesterol was calculated using the Friedewald formula (Friedewald et al. 1972). The iron saturation of transferrin was calculated from the iron and transferrin values as follows:

$$\text{Transferrin saturation} = (\text{FE}/\text{TIBC}) * 100.$$

Total iron-binding capacity (TIBC) was calculated by using the following formula:

$$\text{TIBC} = \text{TRSF} * 22.75.$$

Serum ferritin was determined with a dedicated immunoanalyser (Access, Beckman). 25-Hydroxy vitamin D was determined with an enzyme immunoassay (IDS, OCTEIA^{TM} 25-Hydroxy Vitamin D kit). The ferric-reducing ability of plasma (FRAP) was determined according to the method described by Benzie & Strain (1996). Hydroperoxides in serum were measured as reactive oxygen metabolites (ROM assay) according to the method described by Trotti et al. (2001). The protein thiol oxidation assay (PSH) was based on the method of Hu (1994). The above PSH, FRAP, and ROM assays were adapted for use in micortitre plates.

The intra- and inter-assay per cent coefficients of variation (CVs) for all biomarkers were less than 8% at any time point. Duplicates (two separate aliquots) from the same sample and time point were run either on the same day or after 1 day for the purpose of assay validity and the detection of any systematic error. For each time point (Times 1 and 2), the mean value for these duplicates from a participant was used to represent the biomarker level for that participant. Participants who did not provide enough serum to run duplicates were not included in the assay validity of comparing duplicates at each time point and that single value represented the biomarker level for that participant for the reliability analyses. Only two samples were excluded because of haemolysis.

Data analyses

Paired *t*-tests were used to assess the differences in mean levels of biomarkers at the two times points: Times 1 (T1) and 2 (T2). The individual per cent change between T1 and T2 was also calculated for the biomarkers. The paired *t*-test and per cent change were also used to assess the change of continuous covariates of daily alcohol consumption (ethanol g day^{-1}), daily energy intake (MJ day^{-1}), age (years), and weight (kg). For the categorical variables of current smoking (yes, no), vitamin use (yes, no), and hormone use (yes, no), frequency tables were used for a comparison of

proportions to assess the change in these variables between the two time points. As described above, the Bilthoven population study was composed of males, while the Utrecht population was composed of females. Therefore, the two sites were used to represent categories of gender (i.e. Bilthoven represents male gender and Utrecht represents female gender).

Spearman rank correlation coefficients for non-parametric data, adjusted for the duration of storage of biomarkers, were used to assess the ranking of the biomarkers over time. The intra-class correlation (ICC) was used to assess the inter-individual reliability (relative to the intra-individual reliability) of the biomarkers among individuals over time (Shrout & Fleiss 1979). For the ICC, an unconditional means model using the MIXED procedure of the SAS Institute, Inc. (Cary, NC, USA) was fitted with the study participant as the clustering variable and the repeated values of the biomarkers at T1 and T2 as the dependent variable. ICC values for the total sample were computed unadjustedly and stratified by centre (Bilthoven and Utrecht) as well as smoking status. Further, adjustments for only the duration between the two measurements (in years), and then multivariate adjustment for duration between the two measures, smoking status, age, weight, ethanol intake, energy intake, vitamin use, and hormone use (only for Utrecht) were carried out according to the centres. CRP was log-transformed (to assume normality for parametric tests) in all ICC analyses since it was the only variable that showed substantially different results between the transformed and non-transformed data. All p-values were considered significant at $p < 0.05$. Sample size calculation (using an alpha level of 0.05 and a power of 0.8, and assuming a correlation of 0.85 between measurements at two moments in time) showed that ten participants were enough to show a difference of 20% of the standard deviation of the biomarker. Data from 30 individuals from each centre were collected to take into account the different levels of variability of a broad spectrum of biomarkers.

Results

Most of the analyses carried out were stratified by gender, which identifies the cohorts of Utrecht (women) and Bilthoven (men). In general, most biomarkers did not change significantly between the two time points and changes in mean values were around 10% or less (Table I), which indicates no substantial change of the selected biomarkers over time for the populations. However, there were a few exceptions, such as CRP levels among men was 92% higher at T2 compared with T1, although it was only 5% higher among women. When the CRP was log-transformed, the difference between the two measures among men reduced to 39%. Similarly, when CRP values equal to 10 mg l^{-1}or more were excluded at Time 2 (a CRP higher than 10 mg l^{-1} is an indicator of acute inflammation), the difference was 40%. Other variables did not change appreciably after using the log-transformed data.

There was a statistically significant difference ($p < 0.0001$) in the mean duration between the two points of time for women (2.6 years; range 1.2–4.8) compared with men (5 years, range 4.2–5.9). Age was not statistically different ($p = 0.18$) between the two populations (55 years for women versus 53 years for men).

The levels of the duplicate samples (aliquots) for each biomarker at each time point are well correlated (as shown in Table II), with the exception of PSH and ROM, which were slightly less correlated. This indicates the assay for these two biomarkers has

Table I. Biomarker means and standard errors for Bilthoven and Utrecht, the Netherlands, at times 1 and 2, and the p-value and per cent difference between them.

	Men (Bilthoven)				Women (Utrecht)			
Variable	Time 1, mean (SE)	Time 2, mean (SE)	p	Percentage change	Time 1, mean (SE)	Time 2, mean (SE)	p	Percentage change
Serum ferritin ($\mu g\ l^{-1}$)	116.6 (13.1)	129.8 (13.8)	0.096	11.3	69.6 (11.6)	69.8 (8.5)	0.873	0.2
Total serum iron ($\mu mol\ l^{-1}$)	20.7 (0.80)	22.2 (1.07)	0.264	7.2	18.6 (1.24)	18.2 (0.89)	0.634	−2.2
Transferrin (TRSF) ($g\ l^{-1}$)	2.6 (0.05)	2.8 (0.06)	0.066	7.7	2.8 (0.08)	2.7 (0.07)	0.012	−3.6
Transferrin saturation* (%)	34.9 (1.47)	35.7 (1.95)	0.709	2.3	29.5 (1.85)	29.7 (1.40)	0.862	0.7
Total iron binding capacity (TIBC)* ($\mu mol\ l^{-1}$)	60.0 (1.11)	62.9 (1.42)	0.066	4.8	64.3 (1.85)	62.1 (1.69)	0.012	−2.8
Apo lipoprotein-A ($g\ l^{-1}$)	1.4 (0.03)	1.5 (0.03)	0.302	7.1	1.6 (0.05)	1.5 (0.04)	0.022	−6.3
Apo lipoprotein-B ($g\ l^{-1}$)	1.0 (0.05)	0.9 (0.04)	0.288	−1.0	0.9 (0.03)	0.9 (0.04)	0.972	0.0
Cholesterol ($mmol\ l^{-1}$)	6.0 (0.22)	5.9 (0.21)	0.496	−1.7	6.0 (0.14)	5.9 (0.16)	0.439	−1.7
HDL ($mmol\ l^{-1}$)	1.3 (0.05)	1.4 (0.05)	0.134	7.8	1.6 (0.07)	1.5 (0.06)	0.043	6.3
LDL* ($mmol\ l^{-1}$)	3.9 (0.21)	3.8 (0.19)	0.499	−2.6	3.7 (0.13)	3.7 (0.15)	0.859	0.0
Triglycerides ($mmol\ l^{-1}$)	1.8 (0.13)	1.6 (0.14)	0.145	−11.0	1.6 (0.12)	1.7 (0.19)	0.786	6.3
Uric acid ($\mu mol\ l^{-1}$)	324.6(11.60)	327.4(11.80)	0.668	0.9	278.2 (12.14)	291.2 (11.39)	0.163	4.7
Creatinine ($\mu mol\ l^{-1}$)	80.7 (1.90)	79.5 (1.54)	0.288	−1.5	67.8 (1.49)	72.6 (1.65)	0.001	7.1
Bilirubin ($\mu mol\ l^{-1}$)	8.2 (0.62)	9.4 (0.84)	0.148	14.6	7.0 (0.63)	7.0 (0.65)	0.837	0.0
C-reactive protein ($mg\ l^{-1}$)	2.5 (0.84)	4.8 (2.00)	0.284	92.0	4.0 (0.94)	4.2 (1.06)	0.974	5.0
Fructosamine ($\mu mol\ l^{-1}$)	244.3 (2.78)	252.0 (3.78)	0.005	3.2	241.2 (2.80)	242.2 (3.12)	0.662	0.4
Reactive oxygen metabolites (ROM) ($U\ ml^{-1}$)	360.0 (12.64)	376.1 (11.97)	0.227	4.5	437.8 (18.77)	437.7 (14.43)	0.851	−0.02
Free radical antioxidant potency (FRAP) ($\mu mol\ l^{-1}$)	1059.8 (27.31)	1051.3 (25.62)	0.595	−0.8	963.4 (30.10)	1061.3 (36.98)	0.017	10.2
Protein SH oxidation (PSH) ($\mu mol\ l^{-1}$)	312.5 (4.33)	304.8 (5.43)	0.084	−2.5	296.5 (3.45)	290.3 (4.80)	0.238	−2.1
25-Hydroxy vitamin D ($\mu mol\ l^{-1}$)	77.2 (5.89)	83.5 (5.39)	0.138	8.2	91.2 (6.28)	102.5 (5.58)	0.057	12.4
Alcohol day^{-1} (g)	20.3 (3.73)	22.7 (4.49)	0.287	11.8	7.2 (1.47)	7.8 (1.34)	0.394	8.3
Weight (kg)	81.1 (1.32)	81.1 (1.37)	0.994	0.0	71.5 (2.34)	71.8 (2.29)	0.324	0.4
Energy (kJ)	11017.8 (530.65)	10390.0 (448.17)	0.002	−5.7	1932.7 (85.45)	1836.8 (75.85)	0.009	−4.5
Hormone use (yes, no)					26% (9/35)	40% (14/35)		14
Vitamin use (yes, no)	13% (4/30)	13% (4/30)			31% (11/35)	31% (11/35)		
Smoking (yes, no)	20% (6/30)	20% (6/30)			20% (7/35)	20% (7/35)		

*Computed variables.

Table II. Spearman rank correlation for biomarkers at times 1 and 2 and adjusted correlations for Bilthoven (men) and Utrecht (women).

				Between T1, T2	
Biomarker	Within T1	Within T2	Total	Men*	Women*
Uric acid	0.997	0.996	0.86	0.84	0.86
Serum ferritin	0.997	0.997	0.86	0.75	0.92
HDL	0.997	0.999	0.78	0.70	0.79
Creatinine	0.970	0.932	0.76	0.75	0.78
Apo lipoprotein-A	0.969	0.983	0.69	0.45	0.79
Apo lipoprotein-B	0.992	0.996	0.75	0.69	0.80
LDL	0.997	0.998	0.75	0.65	0.80
25-Hydroxy vitamin D	0.964	0.936	0.65	0.68	0.58
Transferrin (TRSF)	0.807	0.931	0.54	0.22	0.78
Total iron binding capcity (TIBC)	0.809	0.930	0.54	0.22	0.78
Fructosamine	0.938	0.975	0.69	0.75	0.63
Cholesterol	0.995	0.995	0.71	0.63	0.77
Bilirubin	0.974	0.982	0.68	0.43	0.75
Reactive oxygen metabolites (ROM)	0.703	0.578	0.50	0.29	0.39
Protein SH oxidation (PSH)	0.636	0.705	0.53	0.53	0.40
Free radical antioxidant potency (FRAP)	0.933	0.882	0.68	0.78	0.63
Triglycerides	0.998	0.999	0.62	0.64	0.56
Transferrin saturation	0.950	0.989	0.49	0.32	0.57
Total serum iron	0.999	0.994	0.41	0.06	0.56
C-reactive protein (CRP)	0.999	0.999	0.67	0.74	0.54

*Partial Spearman correlation adjusted for duration between the two time points.

some measurement error that could be attenuating the true intra-individual reliability and ranking for these two biomarkers. They were therefore consistently low for the total crude correlation and the gender-specific partial correlation. The highest correlations were observed for ferritin and uric acid ($r = 0.86$) and the cholesterol-related variables, while the lowest was observed for iron-related biomarkers, especially total iron ($r = 0.41$) and transferrin saturation ($r = 0.49$). The latter two were attenuated by the very low correlations among men. In the stratified analyses, the Spearman correlations adjusted for the duration of time between the two measurements were generally higher among women compared with men. The biomarkers of uric acid, ferritin, creatinine, HDL, and apo lipoprotein-B were reliably ranked over time for both men and women. By gender, TRSF (and TIBC, which is calculated from TRSF), total iron, transferrin saturation and apo lipoprotein-A all performed well among women, while they were poorly correlated among men.

Similarly, the biomarkers of uric acid, serum ferritin, creatinine, HDL, and apo lipoprotein-B were the most reliable in the population according to the intraclass correlation (ICC) values; on the other hand, transferrin saturation and total iron were the least reliable (Table III). When the log-transformed values were used, there was no appreciable difference except for the CRP, where ICC values substantially increased. In addition to total crude ICC, Table III shows the crude ICC stratified by smoking status, and by gender, and the latter adjusted for duration between the two measures and for other covariates. In general, adjustment for covariates and duration between the two time points had minimal effect on the ICC. However, ICC values generally increased among smokers compared with non-smokers, with the exception of vitamin

Table III. Crude and adjusted intraclass correlation for each biomarker.

Biomarker	ICC	ICC men	ICC men⋆	ICC men†	ICC women	ICC women⋆	ICC women†	ICC non-smokers	ICC smokers
Uric acid	0.866	0.85	0.86	0.86	0.86	0.86	0.69	0.83	0.96
Serum ferritin	0.841	0.83	0.82	0.75	0.80	0.81	0.80	0.83	0.91
HDL	0.791	0.70	0.67	0.75	0.82	0.79	0.76	0.80	0.74
Creatinine	0.789	0.80	0.81	0.78	0.67	0.67	0.76	0.77	0.85
Apo lipoprotein-A	0.657	0.55	0.49	0.55	0.69	0.67	0.68	0.65	0.73
Apo lipoprotein-B	0.751	0.71	0.72	0.73	0.82	0.82	0.76	0.68	0.95
LDL	0.711	0.64	0.65	0.65	0.83	0.83	0.78	0.63	0.92
25-Hydroxy vitamin D	0.665	0.77	0.78	0.79	0.57	0.58	0.50	0.68	0.45
Transferrin (TRSF)	0.654	0.22	0.20	0.0	0.83	0.83	0.79	0.67	0.45
Total iron binding capcity (TIBC)	0.654	0.22	0.20	0.0	0.83	0.83	0.79	0.67	0.45
Fructosamine	0.649	0.65	0.64	0.67	0.64	0.65	0.62	0.64	0.59
Cholesterol	0.649	0.58	0.59	0.56	0.78	0.78	0.76	0.55	0.94
Bilirubin	0.592	0.35	0.35	0.42	0.82	0.82	0.82	0.52	0.80
Reactive oxygen metabolites (ROM)	0.568	0.43	0.42	0.30	0.53	0.49	0.50	0.54	0.66
Protein SH oxidation (PSH)	0.532	0.60	0.61	0.64	0.38	0.35	0.26	0.57	0.31
Free radical antioxidant potency (FRAP)	0.524	0.83	0.83	0.83	0.38	0.39	0.13	0.40	0.88
Triglycerides	0.523	0.58	0.59	0.64	0.50	0.46	0.47	0.57	0.27
Transferrin saturation	0.491	0.28	0.29	0.28	0.59	0.59	0.51	0.62	0.0
Total serum iron	0.428	0.06	0.08	0.13	0.59	0.60	0.50	0.58	0.0
C-reactive protein (CRP)‡	0.628	0.61	0.62	0.48	0.63	0.62	0.48	0.61	0.70

⋆Adjusted for duration between the two measures.
†Adjusted for weight, age, energy intake, alcohol intake, smoking status, vitamin use, duration between the two time points of sample collection, and hormone use (for women only).
‡Log-transformed.

D, TRSF, TIBC, triglycerides and PSH, while transferrin saturation and total serum iron values were zero among smokers. The results for smokers should be interpreted with caution because of the small number of smokers in our sample (six men and seven women). For ICC by gender, women generally had higher ICC values than men, especially for iron-related variables: TRSF, TIBC, total serum iron, and transferrin saturation. On the other hand, FRAP and PSH values were substantially higher among men. Adjusting for duration between the two times or for other covariates generally attenuated the ICC values, and for men, the TIBC and TRSF reliability reached zero in the multivariate analyses. To explore the apparently low ICC values for iron-related biomarkers among men, we carried out secondary ICC analyses limited to men and found that excluding smokers from the multivariate analyses substantially improved the ICC value for total iron from 0.13 to 0.45, and the transferrin saturation variable (calculated from total iron) from 0.28 to 0.58. However, excluding those with CRP values ≥ 10 mg l^{-1} substantially improved the ICC for TRSF and TIBC (calculated from TRSF) from 0 to 0.53, and serum ferritin from 0.75 to 0.88.

Discussion

The analyses of repeated serum samples from a group of men and women in the Netherlands show that most of the selected biomarkers (serum ferritin, TIBC, TRSF, CRP, bilirubin, cholesterol, triglycerides, apo lipoprotein-A, apo lipoprotein-B, HDL, LDL, uric acid, creatinine, FRAP, fructosamine, and 25-hydroxy vitamin D) were from high to moderate reliability, with few exceptions (total serum iron, transferrin saturation, PSH, ROM) and with some gender differences. Biomarkers in women where generally more reliable than those for men. Uric acid, ferritin, creatinine, HDL, and apo lipoprotein-B levels seemed to be consistently and highly reliable as biomarkers for both men and women, according to the different statistical methods used. In general, the blood lipid-related biomarkers were more consistent over time, followed by markers of oxidative stress and the iron status-related variables. However, iron status-related variables consistently showed low reliability in men, while markers of oxidative stress seemed to be less reliable in women than in men.

Both the inter- and intra-individual reliability of biomarkers can be greatly influenced by storage and handling of the samples prior to analyses, adding another factor of uncertainty to their use for exposure–disease associations in large cohort studies. In general, regardless of the duration, the lower the temperature of deep freezing, the more stable these samples will be (Comstock et al. 1993). The very low temperature of storage at -196°C could explain the stability of the biomarkers in this study over time. Other studies have found that antioxidants such as retinol, alpha-tocopherol, and beta-carotene did not decay after 15 years of deep-freeze storage at -70°C (Comstock et al. 1993). From a sample of 55 Dutch individuals, 25-hydroxy vitamin D was stable and was not degraded after 4 years storage at -20°C (Ocké et al. 1995).

It seems from the analyses that there was no appreciable change in overall levels of our biomarkers of interest over time. CRP among men increased by 92% between Time 1 and Time 2 among men due to very high levels for a few individuals at Time 2 indicating an inflammatory process. Stigant et al. (2005) also reported poor CRP reliability. Some of the low reliability of the biomarkers may be attributed to random error from the laboratory assay. However, this is unlikely to be a major contributor

since all the samples were analysed at the same time period by the same laboratory. Systematic assay error, also known as laboratory drift, is usually addressed by running quality controls in each batch. The lower results for ROM, FRAP, and PSH may be partially attributed to assay variability for these samples because they were performed on two different days. Although quality control samples were run in each batch and results corrected according to these samples, there could still be daily variations due to temperature, preparation of solutions, and other subtle laboratory sources of error.

We were able to assess random error in the laboratory assay by analysing duplicate samples in each of the two time points, and the correlations were very high. The source of variability from external factors was largely adjusted for by only selecting individuals who did not significantly change their diet and vitamin intake, smoking, alcohol consumption, season of blood collection, weight, and had no diseases between the two time points. It is usually difficult to obtain such data at two time points from all participants; therefore, this study addressed it among our selected population. The advantage of including two centres from the same country, as opposed to using two centres from two different countries, is to avoid possible ecological and ethnic differences that can bias the reliability of biomarkers.

There were zero reliability values for TIBC and TRSF for men. This indicates zero intercept variability, and that there was no variation detected for these biomarkers between persons in this sample. In the secondary analyses, we found that the zero values for ICC of TIBC and TRSF among men were due to high CRP values indicative of acute inflammation. TRSF is a negatively-induced acute phase protein that decreases with inflammation, which would explain the poor reliability due to high CRP (Whicher et al. 1991). Smoking on the other hand causes an increase in serum iron levels (Yoshida et al. 2004), which would explain the improvement in the reliability of this biomarker when smokers are excluded in secondary analyses.

A limitation of the study is involving only a Dutch population that might not reflect the true exposure range in the overall population of interest in the EPIC study. However, this is unlikely to effect reliability, although it can affect external validity (we would not extrapolate to all of EPIC). The current study was not fully designed to assess the stability of biomarkers in relation to storage, because we did not have fresh samples from the same individual to compare them with stored samples at different temperatures and for different durations. However, from recent analyses from the same laboratory (unpublished), serum ferritin, total serum iron, apo lipoprotein-A, apo lipoprotein-B, uric acid, creatinine, bilirubin, cholesterol, CRP, HDL, LDL, and triglyceride were analysed as fresh samples, and then stored at -196°C for 1 year and re-analysed. All biomarkers' mean levels were within 95–103% of their baseline mean value, although CRP decreased to 90%, uric acid decreased to 89%, and bilirubin decreased to 85% of their mean values, indicating a possible small storage affect on these three biomarkers. Most of the available literature is focused on the latter approach to assess decay and stability of the biomarkers in relation to storage (Murphy et al. 2000, Wu et al. 2004). One cannot exclude the possibility that the slight differences in the cold chain between samples from Bilthoven retrieved locally and from Utrecht retrieved at IARC may have contributed to the differences observed between men and women in our results.

Conclusion

The present study was able to address the reliability of a wide range of biomarkers that are relevant to many disease outcomes, including cancer, cardiovascular diseases, and diabetes. It was found that serum ferritin was the most reliable biomarker among iron status biomarkers, HDL and apo lipoprotein-B were the most reliable among the blood lipids biomarkers, and uric acid and creatinine were the most reliable among markers of oxidative stress. Biomarkers among women generally had higher reliability than men, and this gender difference needs further investigation.

By using unreliable biomarkers as a one-time exposure in cohort studies, the exposure–disease associations are likely to be lower than the true associations due to external variables and biological variability. Conducting reliability studies before carrying out large and expensive analyses of biomarkers are therefore necessary to identify properly which biomarkers are less likely to cause attenuation of observed exposure–disease associations due to inter- and between-individual variability.

Acknowledgements

The work described in this paper was carried out in part with the financial support of the 'Europe against Cancer' Programme of the European Commission. The Dutch Ministry of Public Health and the National Institute for Public Health and the Environment provided the main funds for the Morgen-EPIC study. The biochemical analyses were made possible by core funding of the National Institute for Public Health and the Environment. The project steering committee of the Morgen-EPIC study consisted of Dr H. B. Bueno-de-Mesquita (PI), Dr H. A. Smit, Dr W. M. M. Verschuren and Dr J. C. Seidell. The Dutch Prevention Fund, LK Research Fund and WCRF provided the main funds for the Prospect-EPIC study. The project steering committee of the Prospect-EPIC study consisted of Dr P. H. M. Peeters (PI) and Dr H. J. A. Collette. The authors would like to acknowledge the work of the laboratory technicians M. E. P. van Hemert-Janssen, L. F. Kregel-van de Velde, D. B. Mooiweer-Boogaerdt, J. C. Heijkoop-Kroon and A. de Wilde for their excellent technical assistance in processing and storing the EPIC blood samples, the laboratory technicians J. W. J. M. Cremers, P. K. Beekhof and H. A. van Loenen for the biochemical analyses, B. J. Slotboom for drawing the Prospect-EPIC sample, and C. Hayes for editorial comments.

References

Al-Delaimy WK, van Kappel AL, Ferrari P, Slimani N, Steghens JP, Bingham S, Johansson I, Wallstrom P, Overvad K, Tjonneland A, Key TJ, Welch AA, Bueno de Mesquita HB, Peeters PH, Boeing H, Linseisen J, Clavel Chapelon F, Guibout C, Navarro C, Quiros JR, Palli D, Celentano E, Trichopoulou A, Benetou V, Kaaks R, Riboli E. 2004. Plasma levels of 6 carotenoids in 9 European countries: report from the European Prospective Investigation into Cancer and Nutrition (EPIC). Public Health Nutrition 6:713–722.

Benzie IFF, Strain JJ. 1996. The ferric-reducing ability of plasma (FRAP) as a measure of 'antioxidant power': The FRAP assay. Annals of Biochemistry 239:70–76.

Bingham S, Riboli E. 2004. Diet and cancer—the European Prospective Investigation into Cancer and Nutrition. Nature Reviews. Cancer 4:206–215.

Boker LK, van Noord PA, van der Schouw YT, Koot NV, Bueno de Mesquita HB, Riboli E, Grobbee DE, Peeters PH. 2001. Prospect-EPIC Utrecht: study design and characteristics of the cohort population. European Prospective Investigation into Cancer and Nutrition. European Journal of Epidemiology 17:1047–1053.

Comstock GW, Alberg AJ, Helzlsouer KJ. 1993. Reported effects of long-term freezer storage on concentrations of retinol, beta-carotene, and alpha-tocopherol in serum or plasma summarized. Clinical Chemistry 39:1075–1078.

De Klerk NH, English DR, Armstrong BK. 1989. A review of the effects of random measurement error on relative risk estimates in epidemiological studies. International Journal of Epidemiology 18:705–712.

De Waard F, Collette HJA, Rombach JJ, Baanders-Van Halewijn EA, Honing C. 1984. The DOM Project for the early detection of breast cancer, Utrecht, the Netherlands. Journal of Chronic Diseases 37:1–44.

Erlinger TP, Platz EA, Rifai N, Helzlsouer KJ. 2004. C-reactive protein and the risk of incident colorectal cancer. Journal of the American Medical Association 291:585–590.

Fracheboud J, de Koning HJ, Beemsterboer PM, Boer R, Hendriks JH, Verbeek AL, van Ineveld BM, de Bruyn AE, van der Maas PJ. 1998. Nation-wide breast cancer screening in the Netherlands: results of initial and subsequent screening 1990–1995. International Journal of Cancer 75:694–698.

Friedewald WT, Levy RI, Fredrickson DS. 1972. Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge. Clinical Chemistry 18:499–502.

Gackowski D, Banaszkiewicz Z, Rozalski R, Jawien A, Olinski R. 2002. Persistent oxidative stress in colorectal carcinoma patients. International Journal of Cancer 101:395–397.

Grant WB, Garland CF. 2004. A critical review of studies on vitamin D in relation to colorectal cancer. Nutrition and Cancer 48:115–123.

Hu ML. 1994. Measurement of protein thiol groups and glutathione in plasma. Methods in Enzymology 233:380–385.

Kato I, Dnistrian AM, Schwartz M, Toniolo P, Koenig K, Shore RE, Zeleniuch Jacquotte A, Akhmedkhanov A, Riboli E. 1999. Iron intake, body iron stores and CRC risk in women: a nested case-control study. International Journal of Cancer 80:693–698.

McKeown-Eyssen GE, Tibshirani R. 1994. Implications of measurement error in exposure for the sample sizes of case-control studies. American Journal of Epidemiology 139:415–421.

Misciagna G, De Michele G, Guerra V, Cisternino AM, Di Leo A, Freudenheim JL, INTEROSP Group. 2004. Serum fructosamine and colorectal adenomas. European Journal of Epidemiology 19:425–432.

Murphy JM, Browne RW, Hill L, Bolelli GF, Abagnato C, Berrino F, Freudenheim J, Trevisan M, Muti P. 2000. Effects of transportation and delay in processing on the stability of nutritional and metabolic biomarkers. Nutrition and Cancer 37:155–160.

Ocké MC, Bueno de Mesquita HB, Goddijn HE, et al. 1997a. The semi-quantitative food frequency questionnaire developed for the EPIC-study in the Netherlands. II. Validity of nutrients. International Journal of Epidemiology 26(Suppl. 1):S49–S58.

Ocké MC, Bueno de Mesquita HB, Jansen AM, Pols MA, Kromhout D, van Staveren WA. 1997b. The semi-quantitative food frequency questionnaire developed for the EPIC-study in the Netherlands. I. Description of questionnaire and validity of food groups. International Journal of Epidemiology 26(Suppl. 1):S37–S48.

Ocké MC, Schrijver J, Obermann-de Boer GL, Bloemberg BP, Haenen GR, Kromhout D. 1995. Stability of blood (pro)vitamins during four years of storage at −20 degrees C: consequences for epidemiologic research. Journal of Clinical Epidemiology 48:1077–1085.

Poynter JN, Gruber SB, Higgins PD, Almog R, Bonner JD, Rennert HS, Low M, Greenson JK, Rennert G. 2005. Statins and the risk of colorectal cancer. New England Journal of Medicine 352:2184–2192.

Riboli E, Hunt KJ, Slimani N, Ferrari P, Norat T, Fahey M, Charrondiere UR, Hemon B, Casagrande C, Vignat J, Overvad K, Tjonneland A, Clavel Chapelon F, Thiebaut A, Wahrendorf J, Boeing H, Trichopoulos D, Trichopoulou A, Vineis P, Palli D, Bueno De Mesquita HB, Peeters PH, Lund E, Engeset D, Gonzalez CA, Barricarte A, Berglund G, Hallmans G, Day NE, Key TJ, Kaaks R, Saracci Rl. 2002. European Prospective Investigation into Cancer and Nutrition (EPIC): study populations and data collection. Public Health Nutrition 5:1113–1124.

Riboli E, Kaaks R. 1997. The EPIC Project: rationale and study design. European Prospective Investigation into Cancer and Nutrition. International Journal of Epidemiology 26:S6–S14.

Shrout PE, Fleiss JL. 1979. Intraclass correlations: uses in assessing rater reliability. Psychological Bulletin 2:420–428.

Smit HA, Verschuren WMM, Bueno-de-Mesquita HB, Seidell JC. 1994. Monitoring van risicofactoren en gezondheid in Nederland (MORGEN-project): Doelstelling en werkwijze. Report No. 263200001. Bilthoven: RIVM.

Stigant CE, Djurdjev O, Levin A. 2005. C-reactive protein levels in patients on maintenance hemodialysis: reliability and reflection on the utility of single measurements. International Urology and Nephrology 37:133–140.

Trotti R, Carratelli M, Barbieri M, Micieli G, Bosone D, Rondanelli M, Bo P. 2001. Oxidative stress and a thrombophilic condition in alcoholics without severe liver disease. Haematologica 86:85–91.

Van Loon AJM, Tijhuis M, Picavet HSJ, Surtees PG, Ormel J. 2003. Survey non-response in the Netherlands: Effects on prevalence estimates and associations. Annals of Epidemiology 13:105–110.

Whicher T, Bienvenu J, Price CP. 1991. Molecular biology, measurement and clinical utility of the acute phase proteins. Pure and Applied Chemistry 63:1111–1116.

Willett WC. 1998. Nutritional epidemiology. 2nd Ed. Oxford: Oxford University Press.

Wu T, Rifai N, Roberts LJ II, Willett WC, Rimm EB. 2004. Stability of measurements of biomarkers of oxidative stress in blood over 36 hours. Cancer Epidemiology and Biomarkers Prevention 13:1399–1402.

Yoshida H, Sasaki K, Hirowatari Y, Kurosawa H, Sato N, Furutani N, Tada N. 2004. Increased serum iron may contribute to enhanced oxidation of low-density lipoprotein in smokers in part through changes in lipoxygenase and catalase. Clinica et Chimica Acta 345:161–170.

Appendix 1. Selection criteria for the participants in the study from Utrecht and Bilthoven at each time point.

Bilthoven & Utrecht	First sample (time 1)	Second sample (time 2)
Number	60 (30 from each center)	60 (30 from each center)
Gender	Males from Bilthoven Females from Utrecht	Males from Bilthoven Females from Utrecht
Age*	40–65 for Bilthoven 50–65 for Utrecht	Any age
Date of collection of samples	Anytime	In the same season of the first sample: within a month of first sample +/− 2 months
Smoking*	Never, former or current	Same category as in the first sample
Alcohol intake*	For Bilthoven: never, former or current For Utrecht: yes or no	Same category as in the first samples
Weight in Kg*	Any weight	Within 5 kg weight of that recorded in the first sample
Energy†	Any energy level	Within the same quartile of energy in the first sample
Vitamin supplements*	Yes or no	Similar to the first sample
Disease history*	No history of diabetes, cardiovascular diseases or cancer	No history of diabetes, cardiovascular diseases or cancer

*general questionnaire.
†FFQ.