Misreporting of energy: prevalence, characteristics of misreporters and influence on observed risk estimates in the Malmö Diet and Cancer cohort

Irene Mattisson*, Elisabet Wirfält, Carin Andrén Aronsson, Peter Wallström, Emily Sonestedt, Bo Gullberg and Göran Berglund

Department of Clinical Sciences, Malmö, Lund University, Malmö University Hospital, SE-205 02 Malmö, Sweden

The present study investigates the prevalence of misreporting of energy in the Malmö Diet and Cancer cohort, and examines anthropometric, socio-economic and lifestyle characteristics of the misreporters. Further, the influence of excluding misreporters on risk estimates of post-menopausal breast cancer was examined. Information of reported energy intake (EI) was obtained from a modified diet history method. A questionnaire provided information on lifestyle and socio-economic characteristics. Individual physical activity level (PAL) was calculated from self-reported information on physical activity at work, leisure time physical activity and household work, and from estimates of hours of sleeping, self-care and passive time. Energy misreporting was defined as having a ratio of EI to BMR outside the 95 % CI limits of the calculated PAL. Logistic regression analysed the risk of being a low-energy reporter or a high-energy reporter. Almost 18 % of the women and 12 % of the men were classified as low-energy reporters, 2·8 % of the women and 3·5 % of the men were classified as high-energy reporters. In both genders high BMI, large waist circumference, short education and being a blue-collar worker were significantly associated with low-energy reporting. High-energy reporting was significantly associated with low BMI, living alone and current smoking. The results add support to the practice of energy adjustment as a means to reduce the influence of errors in risk assessment.

Energy: Misreporting: Physical activity level

Bias in self-reported dietary intakes is a well-known phenomenon; especially the bias in reported energy intake (EI) in relation to energy expenditure (EE) has received attention. Misreporting of energy may be overrepresented in certain population sub-groups and may not be food or nutrient neutral. Such inconsistencies could introduce bias and influence associations between diet and outcome. An extensive review describes characteristics of underreporters and examines problems of data interpretation (Livingstone & Black, 2003). The most robust finding was a positive association between low-energy reporting (LER) and high BMI. The influence of socio-economic status, education, health consciousness, culture and behaviour was less predictable and differed across studies.

A strong negative association between employment grade and prevalence of LER was found in a study examining socio-economic variation in nutrient intakes. For instance in women, top grade had 19·3 % LER and bottom grade 49·2 % LER. Consequently, associations between nutrient intake and socio-economic status changed when underreporters were excluded or when data were energy adjusted (Stallone *et al.* 1997). Desire for weight reduction was found to be associated with underreporting both among normal-weight and overweight persons (Johansson *et al.* 1998).

The association between disease status and diet might be affected by underreporting. Studies on associations between diet, obesity and obesity-related disorders are prone to have interpretation problems (Seidell, 1998). A study addressing differences in the consumption of dietary fat and sugar in men and women with differing BMI (Macdiarmid *et al.* 1998) found that the bias introduced by LER could both create and remove associations between dietary intakes and BMI. This was especially evident in women. For instance, the inclusion of low-energy reporters reversed the association between consumption of high-fat sweet foods and BMI because obese women underreported these products. The associations between diet and the metabolic syndrome were found to be different in underreporters of energy and adequate reporters of energy (Rosell *et al.* 2003). However, a study on relative fat intake and breast cancer risk found no effect on risk estimate when excluding low-energy reporters (Holmes *et al.* 1999).

Misreporting of energy is associated with diverging bias in macronutrient composition. Low-energy reporters had higher percentage of energy from protein (Lilienthal Heitmann & Lissner, 1995; Price *et al.* 1997; Livingstone & Black, 2003) while fat energy tended to be lower (Livingstone & Black, 2003). Similarly, there seems to be a bias in reported foods; low-energy reporters have a tendency to report more 'good' foods like fish, vegetables and fruits and less 'bad' foods like cookies, candies and fat (Becker & Welten, 2001; Livingstone & Black,

2003). Also, food pattern analyses within the Malmö Diet and Cancer (MDC) cohort indicates that clusters of individuals reporting 'healthy foods' had lower mean ratios between energy intake and BMR (Wirfält *et al.* 2000). Because misreporting of dietary intakes may cause erroneous associations and risk estimates, it is important to evaluate EI in epidemiological studies using self-reported dietary data.

The aims of this project are (1) to estimate individual physical activity levels (PAL) within the MDC cohort; (2) to compare different approaches for evaluating the EI; (3) to compare low-energy reporters, adequate energy reporters and high-energy reporters with regard to anthropometry, socio-economic characteristics and lifestyle habits; and (4) to reanalyse the associations between fat, alcohol and fibre and risk of breast cancer in post-menopausal women while excluding misreporters of energy.

Subjects and methods

Malmö Diet and Cancer

The MDC study is a prospective cohort study in Malmö, a city in the south of Sweden with approximately 250 000 inhabitants. In 1991, the MDC source population was defined as all persons living in the City of Malmö and born between 1926 and 1945. In May 1995, the source population was extended to include all women born between 1923 and 1950, and all men born between 1923 and 1945, in total 74 138 individuals. Inadequate Swedish language skills and mental incapacity were the only exclusion criteria. When the baseline examination closed in October 1996, 28 098 participants had complete data sets. Details of the recruitment procedures and the cohort are described elsewhere (Manjer *et al.* 2001). The Ethical Committee at Lund University has approved the MDC study (LU 51–90).

The participants visited the MDC screening centre twice. During the first visit, groups of six to eight participants were instructed in how to register meals in the menu-book (see later) and how to fill out the diet questionnaire and the extensive questionnaire covering socio-economic and lifestyle factors. Project nurses took blood samples, blood pressure and anthropometric measurements. All questionnaires were completed at home. During the second visit, approximately 10 d after the first, the socio-economic questionnaire was checked and the dietary interview conducted.

Study populations

All participants (*n* 28 098) who completed MDC baseline examination were used in the analyses of adequacy of energy reporting. A sub-sample (1706 women and 1188 men) who filled in the first version of the socio-economic questionnaire was used to study the effect of using a constant number of sleeping hours compared to using the individual information about hours of sleeping. A sub-sample of women (*n* 11 726), older than 50 years at baseline, previously described (Mattisson *et al.* 2004*a*), was used in the reanalyses of breast cancer risk in post-menopausal women (Mattisson *et al.* 2004*a,b*).

Diet assessment

The MDC method is an interview-based, modified diet history method. It combines (1) a 7 d menu-book for registration of lunch and dinner meals, cold beverages including alcohol and (2) a questionnaire for assessment of meal pattern, consumption frequencies and portion sizes of regularly eaten foods (i.e. sandwiches, cakes and cookies, fruit, breakfast cereals, milk and yoghurt, coffee and tea, sweets, snacks). Drugs, natural remedies and nutrient supplements were recorded in the menu-book. The participant used a booklet with forty-eight black and white photographs at home to estimate portion sizes in the questionnaire. Usual portion sizes of foods and dishes listed in the menu-book were estimated during the interview from a more extensive book with black and white photographs. Typically, each set of photographs contained four different portion sizes of a dish. In addition, participants were asked complementary questions on their usual meal pattern, cooking methods and details about food choices, for instance, type of fat used in cooking and on bread. Diet interviewers carefully checked the consistency of the information provided in the questionnaire and menu-book. Seventeen trained interviewers performed the interviews. The diet interviewers coded and entered the information from the menu-book during the interview, using interactive software (Kostsvar; AIVO, Stockholm, Sweden).

The mean daily intake of foods was calculated based on frequency and portion size estimates from the questionnaire and menu-book. The food intake was converted to energy and nutrient intakes using the MDC nutrient database where the majority of the nutrient information comes from PC-KOST2-93 from the National Food Administration in Uppsala, Sweden.

The relative validity of the MDC method was evaluated in a sample of Malmö residents, 105 women and 101 men, 50–69 years old, using 18 d of weighed records, 3 d every second month during a year, as the reference method. The Pearson correlation coefficients, between the reference method and the MDC method administrated after the 12-month reference period, were 0·55 for energy in both women and men (Riboli *et al.* 1997).

Variables

Diet. Daily intakes of total energy (EI; MJ/d), non-alcohol energy (MJ/d), fat (g/d), carbohydrates (g/d), protein (g/d) and total alcohol (g/d) were calculated from the dietary information. Further, percentage contribution of fat, carbohydrates and protein to non-alcohol energy was calculated.

In September 1994, the dietary data-processing procedure was slightly altered (Wirfält *et al.* 2002). Method version (indicating data collection before or after 1 September 1994), diet interviewer and season of diet interview were used as covariates to control for undue variation in the dietary data collection over time.

Past change of dietary habits was based on the questionnaire item 'Have you substantially changed your dietary habits because of illness or another reason?'

Anthropometry. Weight was measured to the nearest 0·1 kg using a balance-beam scale with subjects wearing light clothing and no shoes. Standing height was measured with a fixed stadiometer calibrated in centimetres. BMI (kg/m^2) was calculated from direct measurements of weight and height. Relative weight categories (BMI < 18·5; ≥18·5– < 25; ≥25– < 30; ≥30– < 35; ≥35– < 40; ≥40) were used according to current WHO recommendations (World Health Organization Study Group, 2000). However, the two top categories were collapsed into one category because very few individuals (100 women and twenty-eight men) had a BMI ≥ 40. Waist circumference

was measured by a trained nurse midway between the lowest rib margin and iliac crest. Waist was used as a dichotomised variable (women <84 and men <90 in one category and women ≥84 and men ≥94 in another category) according to WHO recommendations (World Health Organization Study Group, 2000). Bioelectric impedance analysis was used for estimating body composition according to procedures provided by the manufacturer (single frequency analyser BIA 103; RLJ-Systems, Detroit, MI, USA). The algorithm supplied by the manufacturer was used to estimate body fat percentage from impedance measurements.

Socio-economic and lifestyle factors. Information on socio-economic and lifestyle factors was collected by a structured multiple-choice questionnaire.

Participants were divided according to their highest level of education: ≤8 years; 9–10 years; 11–12 years; and college education or university degree.

Cohabitant status was assessed by the question 'Do you live alone?' and used as a dichotomous variable with Yes/No alternatives.

Classification of socio-economic status was based on information on job title, tasks and position at work. The procedure was adapted from the one used in the 1989 Swedish population census (National Bureau of Statistics, 1989). In the present study, the socio-economic status information was collapsed into three categories: employers or self-employed, white-collar workers and blue-collar workers. Pensioners were classified according to position before retirement. Housewives (n 592), students (n 94) and unemployed (n 1410) were collapsed into the category 'other'.

The smoking habits of the participants were defined as (1) current smokers, including irregular smokers, (2) ex-smokers or (3) never smokers.

Participants indicated number of minutes per week, separately for the four seasons, for seventeen different leisure-time physical activities (Haftenberger *et al.* 2002). The question was adapted from the Minnesota Leisure Time Physical Activity Questionnaire (Taylor *et al.* 1978; Richardsson *et al.* 1994). For the descriptive parts of the present study, the number of minutes of each activity was multiplied with an activity coefficient and an overall leisure-time physical activity score was created. The score was divided into quintiles and further categorised as low (quintile 1), moderate (quintiles 2–4) or high (quintile 5).

Misreporting of energy. In 1991, Goldberg and Black suggested the use of fundamental principles of energy physiology for the estimation of EE to evaluate the level of EI in self-report dietary data (Goldberg *et al.* 1991). According to recommendations from WHO, total EE can be expressed as multiples of BMR, the so-called PAL. WHO provided guidelines for the estimation of PAL based on physical activity and algorithms, based on gender, age, weight and height, for the calculation of BMR (FAO/WHO/UNU, 1985). EI can also be expressed as multiples of calculated BMR and evaluated against the estimated PAL (Goldberg *et al.* 1991).

The procedures for evaluating EI have been refined since 1991. Originally, a fixed PAL was suggested to evaluate energy intakes at both group level and individual level (Goldberg *et al.* 1991). However, in 2000, Black emphasised the need to use individually estimated PAL based on the information on physical activity during leisure time and work. Black also pointed out that the so-called 'Goldberg cut-off point 1' from 1991 should be used only at group level and not at the individual level (Black, 2000*a,b*).

BMR. BMR was estimated with the Schofield equation (FAO/WHO/UNU, 1985) and based on information on age, sex, height and weight of the participants.

Total physical activity level (PAL_tot). In order to estimate individual PAL, the 24 h were divided into time segments of different activities for each participant. Intensity factors (the ratio of work metabolic rate to resting metabolic rate; MET) were assigned to each specific activity (Ainsworth *et al.* 1993). We used information from the questionnaire concerning hours and intensity of leisure-time physical activity, hours of household work, and hours and intensity of occupational activities. Hours of sleeping, time for self-care and 'passive time' were estimated.

Information about leisure-time physical activity from the questionnaire (see earlier) was used to estimate PAL from leisure time. First, the numbers of minutes per week of each activity were converted to hours per day, and each activity was assigned an intensity (MET) factor (Ainsworth *et al.* 1993; (see Appendix 1). Secondly, PAL was calculated for each specific activity as: hours per day $\times$ MET factor/24. PAL from leisure-time physical activity (PAL_ltpa) was calculated for each participant as the sum of PAL from all activities.

The questionnaire included questions on present occupation, hours of work per week and physical intensity at work. For participants gainfully employed at baseline we calculated the number of hours of work per day (hours per week/7) and the different intensity levels of occupational activities were assigned an activity factor (Ainsworth *et al.* 1993; see Appendix 2). For each participant PAL_work was calculated as: hours of work per day $\times$ MET factor/24.

Information on hours of household work was used to estimate PAL from household work (PAL_house). A MET factor of 2·5 (Ainsworth *et al.* 1993) was assigned to household work and PAL_house was calculated as: hours of household work per day $\times$ 2·5/24.

Different versions of the socio-economic questionnaire were used during baseline examination. Because of this, the information on individual sleeping hours was available only in a sub-sample of 1706 women and 1188 men. The calculated median time for sleeping was 7.3 h in all gender and age sub-groups in the this sub-sample. Thus, 7.3 h sleeping were used in all individuals. A MET factor of 0·9 (Ainsworth *et al.* 1993) was assigned to sleeping and PAL_sleep was calculated as: $7·3 \times 0·9/24$.

PAL from self-care (PAL_self) was constant for all participants. We allocated 1 h/d for self-care and used a MET factor of 1·6 (Ainsworth *et al.* 1993) for these activities. PAL_self was calculated as: $1 \times 1·6/24$.

Passive time was calculated for each participant by subtracting 'active time' (i.e. hours of leisure-time physical activity, hours of work, hours of household work, 1 h self-care and 7.3 h sleeping) from the 24 h. 'Active time' exceeded 24 h in seven men and in twenty-one women, and was equal to 8.3 h (i.e. no leisure-time physical activity, work or household work was reported) in forty-five men and in forty women. In those with active time exceeding 24 h, the number of hours of leisure-time physical activity, hours of work and hours of household work was adjusted, while hours of self-care and sleeping remained constant. A MET factor of 1·3 (Ainsworth *et al.* 1993) was assigned to passive-time activities. PAL_pass was calculated as: hours of passive time $\times$ 1·3/24.

Total PAL (PAL_tot) was calculated as the sum of PAL_ltpa, PAL_work, PAL_house, PAL_sleep, PAL_self and PAL_pass.

Classification of energy misreporters. For each participant the adequacy of energy reporting was defined in three different ways in line with current routines in epidemiological studies.

Firstly, the confidence limits for the agreement between EI:BMR and individual PAL were calculated according to recommendations for evaluating 'habitual' intake in individuals by Black (2000*a*). The 95 % confidence limits were calculated according to the following equation.

$$\exp[\pm 2(S/100)], \quad \text{where} \quad S = \sqrt{(CV_{wB}^2 + CV_{tP}^2)}$$

CV_{wB} is the coefficient of variation of the within-person variation for predictions of BMR using the Schofield equation. We used the coefficients of variation suggested for different age and gender groups by Black (2000*a*). CV_{tP} is the coefficients of variation of the total variation in PAL. We estimated these coefficients of variation in the sub-sample with individual information on sleeping hours (Table 1). A CV_{tP} of 15 % is suggested as an average substitute value by Black (2000*a*), thus coefficients of variation used in the present study are higher.

Individuals with EI:BMR below the lower 95 % confidence limit were classified as low-energy reporters. Individuals with EI/BMR within the confidence limits were classified as adequate-energy reporters and individuals with EI:BMR above the upper 95 % confidence limit were classified as high-energy reporters.

Secondly, EI:BMR was compared with fixed cut-off points for PAL (Goldberg *et al.* 1991; Goldberg & Black, 1998). LER EI:BMR ≤ 1·35, AER 1·35 < EI:BMR ≤ 2·82 and HER EI:BMR > 2·82.

Thirdly, EI:BMR was ranked and the top 1 % categorised as HER and the bottom 1 % categorised as LER. This is according to principles used by the European Prospective Investigation into Cancer and Nutrition (Bingham *et al.* 2003).

Statistical methods

SPSS statistical computer package version 10.0 (SPSS Inc., Chicago, IL, USA) was used for the statistical analyses.

The effect of using an estimate for the number of sleeping hours compared to individual information was tested in a sub-sample with complete information. In the first version of the questionnaire (1706 women and 1188 men) the individual information about hours of sleep was used and total PAL (PALi) was estimated. In a second step, the individual information was replaced with the median hours of sleep (7·3 h in all gender and age subgroups) and total PAL (PALc) was calculated. Mean differences (PALi − PALc) were calculated.

The categorisation of participants into LER, AER and HER using individual PAL was compared with the two other procedures of evaluating energy reporting by cross-classification of categories.

Mean differences of baseline characteristics between LER, AER and HER were tested with one-way ANOVA. χ^2 tests were used to test differences in proportions across categories.

The risk of being classified as LER rather than AER, and of being classified as HER rather than AER, was estimated with logistic regression. First, the risk associated with BMI, waist, education, cohabitation, socio-economic status and smoking was estimated while adjusting for age, diet interviewer, season of interview, method version and change of dietary habits. Secondly, all variables were entered simultaneously into the regression to assess the genuine effect on risk.

Previously we have shown a non-significant increased risk for breast cancer in post-menopausal women with high alcohol intakes (Mattisson *et al.* 2004*b*); a significant trend across quintiles of relative fat (Mattisson *et al.* 2004*b*) and a protective effect of high fibre intakes on breast cancer risk (Mattisson *et al.* 2004*a*). Cox regression was used to estimate the association between intakes of relative fat, alcohol, energy-adjusted fibre intakes and risk of breast cancer in post-menopausal women while excluding misreporters of energy. All variables and procedures are described in detail elsewhere (Mattisson *et al.* 2004*a,b*).

Results

Median difference between PALi and PALc was −0·039 in men and −0·037 in women. Differences ranged between −0·123 and 0·036 in men, and −0·102 and 0·031 in women. In women, 1·6 % (twenty-one out of 1310 women) of those falling in the AER category with PALi were classified as low-energy reporters with PALc and 15 % (twelve out of seventy-nine women) of those classified as high-energy reporters using PALi were classified as adequate reporters using PALc. Corresponding figures in men were 1·2 % (twelve out of 989 men) and 22 % (thirteen out of fifty-eight men).

Baseline characteristics are shown in Tables 2 and 3. In women, 17·6 % were classified as LER compared with 12·1 % in men. A slightly higher proportion of men were high-energy reporters, 3·5 % compared with 2·8 % in women. Bivariate analyses show that low-energy reporters of both genders had lowest age, highest weight, highest BMI, highest body fat percentage and largest waist circumference. Estimated BMR and PAL were highest in low-energy reporters. In addition, nutrient density differed; low-energy reporters had highest energy percentage from protein and carbohydrates and lowest from fat. Alcohol intakes were highest in adequate-energy reporters. The highest proportion of current smokers and people living alone was found in the HER category. Low-energy reporters had the highest proportion of high leisure-time physical activity, blue-collar workers and low education.

There were large discrepancies in the classification into different energy-reporting categories depending on the definition of misreporting of energy (Table 4). About one-third (women 35·0 % and men 33·2 %) of those classified as AER when using individual PAL were classified as underreporters when using a fixed cut-off point for EI:BMR. About 85 % (women 87·3 % and men 84·9 %) of the HER category, defined with individual PAL, were classified as reporting adequate energy when using a fixed EI:BMR cut-off. The exclusion of the bottom and top 1 %

Table 1. CV of physical activity level estimated from a sub-sample of the Malmö Diet and Cancer cohort

Age groups	Women		Men	
	n	CV_{tP}	*n*	CV_{tP}
≤ 60 years	1110	19·4	692	21·3
> 60 years	606	17·9	496	19·5

CV_{tP}, coefficients of variation of the total variation in PAL.

Table 2. Differences in baseline characteristics across categories of energy reporting in the Malmö Diet and Cancer cohort, 1991–6

Women

	All (n 17 009)		Low-energy reporters (n 2998; 17·6%)		Adequate-energy reporters (n 13 532; 79·6%)		High-energy reporters (n 479; 2·8%)		F test*
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	P value
Age	57·4	7·9	55·4[a]	7·1	57·8[b]	8·0	59·1[c]	7·9	<0·01
Height (cm)	164	6·1	164[a]	6·0	164[a]	6·1	162[b]	6·4	<0·01
Weight (kg)	68·0	11·7	72·4[a]	13·1	67·3[b]	11·1	61·0[c]	10·5	<0·01
BMI	25·4	4·2	27·1[a]	4·7	25·1[b]	4·0	23·1[c]	3·9	<0·01
Body fat (%)	30·8	5·0	32·4[a]	4·9	30·5[b]	4·9	28·0[c]	5·4	<0·01
Waist (cm)	77·9	10·6	81·2[a]	11·7	77·3[b]	10·2	73·5[c]	9·2	<0·01
Energy intake (MJ/d)	8·4	2·1	6·3[a]	1·3	8·6[b]	1·8	13·3[c]	2·1	<0·01
BMR (MJ/d) (Schofield)	5·8	0·48	6·0[a]	0·50	5·8[b]	0·45	5·5[c]	0·46	<0·01
EI/BMR	1·46	0·38	1·05[a]	0·22	1·50[b]	0·31	2·44[c]	0·37	<0·01
PAL	1·62	0·31	1·91[a]	0·38	1·57[b]	0·26	1·42[c]	0·14	<0·01
Alcohol (g/d)†	7·7	10·6	3·6[a]	9·0	5·7[b]	11·0	3·8[a]	11·0	<0·01
Fibre (g/d)†	18·1	7·8	15·3[a]	6·7	18·5[b]	7·5	25·2[c]	10·3	<0·01
Energy from fat (%)‡	38·5	6·0	36·4[a]	6·0	38·8[b]	5·9	41·9[c]	6·0	<0·01
Energy from carbohydrates (%)‡	45·5	6·0	46·3[a]	6·1	45·4[b]	5·9	44·2[c]	5·9	<0·01
Energy from protein (%)‡	16·0	2·8	17·2[a]	2·8	15·8[b]	2·4	14·0[c]	2·1	<0·01

Men

	All (n 11 047)		Low-energy reporters (n 1333; 12·1%)		Adequate-energy reporters (n 9324; 84·4%)		High-energy reporters (n 390; 3·5%)		F test*
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	P value
Age	59·2	7·1	55·9[a]	5·8	59·8[b]	7·1	61·9[c]	6·5	<0·01
Height (cm)	176	6·6	177[a]	6·7	176[a]	6·6	174[b]	6·7	<0·01
Weight (kg)	81·7	12·1	86·3[a]	12·7	81·4[b]	11·8	74·5[c]	12·2	<0·01
BMI	26·2	3·5	27·6[a]	3·5	26·1[b]	3·4	24·5[c]	3·7	<0·01
Body fat (%)	20·8	5·0	21·9[a]	4·9	20·6[b]	5·00	19·4[c]	5·6	<0·01
Waist (cm)	93·7	10·0	97·1[a]	10·3	93·4[b]	9·9	89·3[c]	10·4	<0·01
Energy intake (MJ/d)	10·8	2·8	8·4[a]	1·9	11·0[b]	2·5	16·5[c]	3·0	<0·01
BMR (MJ/d) (Schofield)	7·3	0·79	7·8[a]	0·69	7·3[b]	0·76	6·7[c]	0·78	<0·01
EI/BMR	1·50	0·41	1·08[a]	0·24	1·52[b]	0·34	2·45[c]	0·41	<0·01
PAL	1·60	0·38	2·07[a]	0·45	1·55[b]	0·32	1·35[c]	0·16	<0·01
Alcohol (g/d)†	11·4	18·5	9·3[a]	16·5	11·8[b]	18·6	9·5[a]	19·5	<0·01
Fibre (g/d)†	20·5	9·4	17·3[a]	7·9	20·8[b]	9·1	28·0[c]	11·3	<0·01
Energy from fat (%)‡	39·8	6·2	38·1[a]	6·2	39·9[b]	6·2	42·8[c]	6·4	<0·01
Energy from carbohydrates (%)‡	44·7	6·2	45·2[a]	6·3	44·7[a]	6·1	43·3[b]	6·3	<0·01
Energy from protein (%)‡	15·5	2·5	16·7[a]	2·8	15·4[b]	2·4	13·9[c]	2·3	<0·01

EI, reported total energy intake; PAL, physical activity level.

[a,b,c] Mean values within a row with the same letters indicate homogenous subsets with Tukey *post hoc* test.

* One-way ANOVA.

† Median (inter-quartile range) intake reported, tests made on geometric means.

‡ Based on non-alcohol energy.

Table 3. Differences in baseline characteristics (percentages) across categories of energy reporting in the Malmö Diet and Cancer cohort 1991–6

	Women					Men				
	All	Low-energy reporters	Adequate-energy reporters	High-energy reporters	χ² P value	All	Low-energy reporters	Adequate-energy reporters	High-energy reporters	χ² P value
Smoking habits					<0·01					<0·01
Current smoker	28·1	30·2	27·2	39·8		28·6	28·0	28·1	41·0	
Ex-smoker	27·7	28·1	27·9	20·4		43·2	43·7	43·5	34·4	
Never smoker	44·2	41·7	44·9	39·8		28·2	28·3	28·4	24·6	
Leisure-time physical activity					<0·01					<0·01
Low	19·6	18·7	19·7	24·0		20·8	22·3	20·1	30·7	
Medium	61·6	56·1	62·7	63·9		57·3	53·6	57·9	53·9	
High	18·8	25·2	17·6	12·0		22·0	24·1	22·0	15·4	
Do you live alone?					<0·01					<0·01
Yes	28·2	24·2	28·7	40·1		18·7	19·4	18·2	29·1	
No	71·8	75·8	71·3	59·9		81·3	80·6	81·8	70·9	
Socio-economic status					<0·01					<0·01
Blue-collar worker	31·2	53·1	25·0	16·8		30·8	48·2	26·9	33·0	
White-collar worker	60·7	39·6	66·8	73·1		49·3	26·9	53·2	50·0	
Self-employed or employers	8·1	7·3	8·2	10·2		19·9	19·9	20·0	17·0	
Educational level					<0·01					<0·01
≤ 8 years	39·4	46·3	37·8	42·4		46·0	55·0	44·5	51·2	
9–10 years	30·4	25·9	31·4	28·5		19·7	20·6	19·6	18·3	
11–12 years	7·0	6·0	7·2	7·8		11·9	9·2	12·3	10·6	
College or university	23·2	21·8	23·6	21·3		22·5	15·2	23·6	19·9	

Table 4. Agreement in classification (percentages) between different procedures for the categorisation of reported total energy intake (EI)

	Individual physical activity level		
	Low-energy reporters	Adequate-energy reporters	High-energy reporters
Fixed cut-off for EI:BMR			
Women			
Low-energy reporters	90·9	35·0	0·0
Adequate-energy reporters	9·1	64·9	87·3
High-energy reporters	0·0	0·1	12·7
Men			
Low-energy reporters	87·4	33·2	0·2
Adequate-energy reporters	12·6	66·5	84·9
High-energy reporters	0·0	0·3	14·9
EI:BMR distribution			
Women			
Low-energy reporters	6·5	0·0	0·0
Adequate energy reporters	93·5	99·8	79·3
High-energy reporters	0·0	0·2	20·7
Men			
Low-energy reporters	6·3	0·0	0·0
Adequate-energy reporters	93·7	99·4	74·9
High-energy reporters	0·0	0·6	25·1

of the ratio EI:BMR only identified a sub-fraction of misreporters of energy (see Table 4).

In multivariate analysis (models adjusted for age, diet interviewer, season of interview, method version and change of dietary habits; data not shown), high BMI, large waist circumference, short education and being a blue-collar worker were significantly associated with higher risk of being a low-energy reporter in both genders. There were no significant associations with cohabitant status or smoking habits. Most significant associations remained in the analysis including mutual adjustment. However, there

Table 5. Multivariate adjusted* risk† of being a low-energy reporter compared to being an adequate-energy reporter in the Malmö Diet and Cancer study

	Men			Women		
	Mean	95 % CI	*n*	Mean	95 % CI	*n*
BMI						
< 18·5	0·31	0·42, 2·30	45	0·57	0·34, 0·95	213
≥ 18·5− < 25	1		3836	1		8333
≥ 25− < 30	1·81	1·52, 2·15	5304	1·64	1·47, 1·84	5508
≥ 30− < 35	2·92	2·29, 3·71	1190	2·45	2·08, 2·89	1749
≤ 35	3·19	2·13, 4·76	183	3·58	2·84, 4·51	493
Waist						
< 94/80 cm	1		5468	1		10 234
≥ 94/80 cm	1·23	1·04, 1·44	5090	1·23	1·09, 1·38	6062
Education						
≤ 8 years	1		4828	1		6376
9−10 years	0·85	0·72, 1·01	2081	0·85	0·76, 0·95	4977
11−12 years	0·67	0·53, 0·83	1257	0·75	0·62, 0·91	1113
College or university	0·57	0·47, 0·69	2392	1·09	0·95, 1·24	3810
Socio-economic status						
Blue-collar worker	1		3483	1		5699
White-collar worker	0·52	0·45, 0·61	4737	0·38	0·35, 0·43	8203
Self-employed/employers	0·86	0·72, 1·03	1722	0·56	0·46, 0·66	1070
Do you live alone?						
No	1		8630	1		11 763
Yes	1·05	0·90, 1·23	1928	0·90	0·82, 1·00	4533
Smoking						
Never smoker	1		2993	1		7212
Ex-smoker	0·92	0·80, 1·07	4599	1·07	0·97, 1·19	4560
Current	0·84	0·72, 0·99	2966	1·12	1·01, 1·24	4524

* Adjusted for age, season of diet interview, diet interviewer, method version, change of dietary habits and all variables in the Table.
† Odds ratio (95 % CI)·

Table 6. Incidence rate ratio (IRR)* of breast cancer according to intake categories of total alcohol in the Malmö Diet and Cancer cohort 1991–2001

Total alcohol intake (g/d)	Adequate-energy reporters only				All women†			
	Person-years	Cases	IRR	95 % CI	Person-years	Cases	IRR	95 % CI
Abstainers	5340	15	0·83	0·49, 1·42	7231	22	0·89	0·57, 1·39
≤15	53 424	205	1		67 244	257	1	
>15−30	9179	31	0·81	0·55, 1·19	10 663	39	0·88	0·62, 1·24
>30	1458	11	1·86	1·00, 3·47	1637	11	1·69	0·91, 3·12

* Adjusted for diet interviewer, method version, season of diet interview, age at baseline, total energy, change of dietary habits, height, waist, current hormone use, age at first child, age at menarche, leisure-time physical activity, smoking habits, educational level.
† Data from Mattisson *et al.* (2004*b*).

was no longer any difference between college or university education and short education in women (see Table 5).

Higher risk of being classified as HER was, in mutually adjusted multivariate models, associated with low BMI, small waist circumference (men only), living alone and current smoking (data not shown).

In order to compare with previously published reports (Mattisson *et al.* 2004*a,b*), Cox regression analyses were used to assess the association between alcohol, fat and fibre and risk of breast cancer in post-menopausal women, excluding those reporting inadequate energy intakes. The risk associated with high intakes of alcohol was slightly higher than the previously reported risk estimates (see Table 6). *P* for positive trend across quintiles of relative fat intake changed from 0·018 to 0·021 and relative risk in highest quintile of fibre intake changed from 0·58 (95 % CI 0·40, 0·84) in all women to 0·59 (95 % CI 0·39, 0·89) in adequate-energy reporters.

Discussion

Approximately 18 % of the women and 12 % of the men in the MDC cohort were classified as LER when EI:BMR was compared with the 95 % confidence limits of the individually estimated PAL. Classification into LER, AER and HER categories according to this procedure differed substantially compared with using a fixed cut-off point for PAL. Higher risks of being a low-energy reporter were, in both genders, associated with high BMI, large waist, short education and being a blue-collar worker. Low BMI, living alone and current smoking were, in both gender groups, associated with higher risk of being a high-energy reporter. In the Cox regression models including only adequate-energy reporters, high alcohol intakes were associated with significantly increased risk of post-menopausal breast cancer. Thus, in comparison with analyses including all women (Mattisson *et al.* 2004*b*), risk estimates were higher. The associations between energy-adjusted fibre intake or relative fat intake and risk of breast cancer when excluding misreporters of energy were similar to those obtained when all women were included (Mattisson *et al.* 2004*a*).

The procedures used in the present study are based on self-reports which are prone to misreporting and results could be influenced for instance by correlated biases between underreporting energy intake and overreporting physical activity. When comparing with the routines used in large-scale epidemiological studies to classify misreporting with our approach, large differences were noted. The use of a fixed cut-off point (PAL=1·35) overestimated LER in both men and women and severely underestimated HER compared to using individual PAL. It is known that LER is more common than

HER (Black, 2000*b*). Excluding the top and bottom 1 % of the distribution of EI:BMR captures only a small fraction (about 6 % in both genders) of those classified as low-energy reporters with individual PAL. HER is also underestimated when this procedure is compared to the individual PAL procedure.

In the Harvard cohorts, participants are sometimes excluded based on absolute energy intakes. For instance, in women Hu *et al.* (1997) used a lower limit of <500 kcal/d and an upper limit of >3500 kcal/d. Salmerón *et al.* (1997) used a lower limit in men of <800 kcal/d and an upper limit of 4200 kcal/d. These cut-offs would not identify any underreporting in the MDC cohort because no woman and only two men were below the lower limit. In the MDC cohort 246 men and 150 women had energy intakes higher than the upper limits. Of those classified as HER in the MDC cohort approximately 20 % of the women and 30 % of the men were identified above the upper limits. However, different diet assessment methods are used and these comparisons indicate that the estimates of absolute levels of energy intake vary with methodology.

Comparisons of the prevalence of misreporting of energy between studies are difficult because of the different criteria used to classify LER and HER. A study in Swedish men (Rosell *et al.* 2003), using similar procedures for classifying LER as the present study, found that 29 % of men were classified as underreporters (all others were classified as non-underreporters and not categorised into adequate or overreporting). However, Rosell *et al.* (2003) used a 7 d food record as the diet assessment instrument.

Recent analyses of data from the validation study of the MDC method (V Kipnis, I Mattisson, D Midthune *et al.*, unpublished results) show that mean daily intake of protein estimated from urinary N and the MDC method were close. Protein estimated from urinary N was 72·4 g in women and 88·0 g in men; protein from diet was 74·5 g in women and 90·7 g in men. This indicates that prevalence of LER might be lower using the MDC method compared to record methods.

Misreporters of energy were excluded from the re-examinations of breast cancer risk associated with intakes of fat, fibre and alcohol. The breast cancer risk associated with high absolute alcohol intake was higher while the other risk estimates were largely unchanged. Misreporting of energy is not necessarily the same as misreporting of alcohol since macronutrients might be selectively misreported (Subar *et al.* 2003). On the other hand, alcohol is notoriously underreported and this could contribute to the misreporting of total EI. Alcohol intake categories were based on absolute intake levels with energy included in the model (standard multivariate model) while fat and fibre intake categories were defined as quintiles of energy-adjusted intakes (residual method). One effect of energy adjustment is to reduce the

influence of errors (Willett *et al.* 1997) in dietary assessment and the residual technique appears more efficient than the standard multivariate model (Kipnis *et al.* 2003). This could explain why there were no differences in outcome for fat and fibre when misreporters of energy were excluded from the risk analyses. Further, this supports the necessity of energy adjustment when analysing self-reported dietary intakes (Kipnis *et al.* 2003). Using absolute values, of nutrient or of energy intakes, as exposure variables might lead to erroneous conclusions.

The most robust finding, also when using doubly labelled water for the estimation of EE (Tooze *et al.* 2004), in studies exploring characteristics of misreporters of energy is the inverse association with BMI. Interestingly, there is no threshold in the risk of underreporting across BMI categories but a trend from the lowest to the highest BMI category (see Table 5). This indicates that part of the effect of BMI might be due to 'intake-related bias' ('scaling bias', Kipnis *et al.* 1999, 2003), i.e. higher energy intakes are underreported and lower intakes are overreported.

A study comparing different procedures to estimate EE in women concluded that the WHO procedure might give biased results with respect to body fatness of the subjects (Löf *et al.* 2003). Predicted BMR tended to be too high when body fatness increased, while the MET factors tended to be too low with increasing body fatness. Thus, using the WHO procedures might lead to an overestimation of underreporting of energy in the obese in the present study.

Today, the only way to obtain unbiased information on EE is to use doubly labelled water as a biomarker. This technique is very expensive and not possible to use in large-scale epidemiological studies. Thus, alternative procedures for the estimation of EE are used. In the present study, we estimated individual PAL values and used these for the relative validation of EI according to Black (2000*a*). All of these procedures have limitations. For instance, the same MET values are used for all individuals regardless of age, gender and body size. One has to consider very carefully if associations found between underreporting and other characteristics are true, or artefacts caused by these procedures.

In the present study, age differed between low-, adequate- and high-energy reporters (see Table 1 and 2). Most of the studies using single cut-off for PAL for the classification of reported EI have found a higher proportion of LER among older subjects. However, studies with measured EE are inconsistent in the association between age and underreporting.

The risk of HER was associated with current smoking. Smoking is associated with reduced food intake and lower body weight (Young-Hwan *et al.* 2002). However, nicotine also increases EE (Hofstetter *et al.* 1986) and using published energy costs of activities might underestimate EE in smokers (Warwick & Busby, 2004). The higher risk of HER in smokers could thus be an effect of underestimated EE and not HER. However, in women, an increased risk of LER was also found in smokers. This finding might indicate that female smokers misreport their dietary intake unpredictably, some overreport and some underreport.

We estimated PAL based on self-reported information on hours and intensity of leisure-time physical activity, hours of household work, and hours and intensity of occupational activities. As for all self-reported data there is a risk of misreporting and bias that might influence results.

We did not have any individual information on hours of self-care and on 'passive time' which is a limitation. In addition,

information on sleeping hours was not available for the entire cohort and thus had to be estimated from a sub-sample. This led to a higher mean PAL_tot and the proportion of HER was underestimated compared to using individual sleeping hours. However, when baseline characteristics of LER, AER and HER of the sub-sample with individual information on sleeping hours were compared with the entire cohort, patterns were in general the same (data not shown).

In addition, the contribution from PAL_work to the total individual PAL is substantial. In *ad hoc* analyses of activity at baseline we found that the multivariate adjusted risk of being a low-energy reporter was very low for the categories 'pensioners' (in men: relative risk 0·17; 95 % CI 0·14, 0·22; in women: relative risk 0·29; 95 % CI 0·25, 0·33) and 'housewife, unemployed etc.' (in men: relative risk 0·15; 95 % CI 0·10, 0·21; in women: relative risk 0·16; 95 % CI 0·13, 0·20) compared with 'gainfully employed'. Thus, using a constant time for sleeping and self-care might introduce larger biases in the estimated PAL in these categories. This could also contribute to the result that the mean age was lower in LER (Table 2) since 'pensioners' have higher age than 'gainfully employed'.

To conclude, risk of LER was associated with high BMI, large waist circumference, short education and being a blue-collar worker in both genders and with current smoking and not living alone in women only. Risk of HER was associated with low BMI, small waist (men only), living alone and current smoking. However, these associations could be influenced by, or be artifacts of, the procedures used for estimating physical activity, EE and BMR. Studies using unbiased measurements of EE and BMR are therefore needed to better elucidate the characteristics of misreporters of energy. When misreporters of energy were excluded from the analyses, the previously observed association between high absolute alcohol intake and post-menopausal breast cancer risk was strengthened, while associations with energy-adjusted intakes of fat and fibre remained largely unchanged. The observations support the use of energy adjustment in analysis to reduce the influence of errors in risk assessment.

References

Ainsworth BE, Haskell WL, Leon AS, Jacobs DR, Montage HJ, Sallis JF & Paffenbarger RS (1993) Compendium of physical activity: classification of energy costs of human physical activities. *Med Sci Sports Exerc* **25**, 71–80.

Becker W & Welten D (2001) Under-reporting in dietary surveys – implications for development of food-based dietary guidelines. *Public Health Nutr* **4**, 683–687.

Bingham SA, Day NE, Luben R, et al. (2003) Dietary fibre in food and protection against colorectal cancer in the European Prospective Investigation into Cancer and Nutrition (EPIC): an observational study. *Lancet* **361**, 1496–1501.

Black AE (2000*a*) Critical evaluation of energy intake using the Goldberg cut-off for energy intake:basal metabolic rate. A practical guide to its calculation, use and limitations. *Int J Obes* **24**, 1119–1130.

Black AE (2000*b*) The sensitivity and specificity of the Goldberg cut-off for EI:BMR for identifying diet reports of poor validity. *Eur J Clin Nutr* **54**, 395–404.

FAO/WHO/UNU (1985) *Energy and Protein Requirements. Report of a Joint FAO/WHO/UNU Expert Consultation.* Technical Reports Series no. 724. Geneva: WHO.

Goldberg GR & Black AE (1998) Assessment of the validity of reported energy intakes – review and recent development. *Scand J Nutr* **42**, 6–9.

Goldberg GR, Black AE, Jebb SA, Cole TJ, Murgatroyd PR, Coward WA & Prentice AM (1991) Critical evaluation of energy intake using fundamental principles of energy physiology: 1 Derivation of cut-off limits to identify under-recording. *Eur J Clin Nutr* **45**, 569–581.

Haftenberger M, Schuit AJ, Tormo MJ, Boeing H, Wareham N, Bueno-de-Mesgiuta Bas H, Kumle M, Hjartåker, Chirlaque MD & Ardanaz E (2002) Physical activity of subjects aged 50-64 years involved in the European Prospective Investigation into Cancer and Nutrition (EPIC). *Public Health Nutr* **5**, 1163–1177.

Hofstetter A, Schutz Y, Jequier E & Wahren J (1986) Increased 24-hour energy expenditure in cigarette smokers. *N Engl J Med* **314**, 79–82.

Holmes MD, Hunter DJ, Colditz GA, Stampfer MJ, Harkinsson SE, Speizer FE, Rosner BA & Willett WC (1999) Association of dietary intake of fat and fatty acids with risk of breast cancer. *JAMA* **281**, 914–920.

Hu FB, Stampfer MJ, Manson JE, Rimm E, Colditz GA, Rosner BA, Hennekens CH & Willett WC (1997) Dietary fat intake and the risk of coronary heart disease in women. *N Engl J Med* **337**, 1491–1499.

Johansson L, Solvoll K, Bjorneboe G-Eaa & Drevon CA (1998) Under- and overreporting of energy intake related to weight status and lifestyle in a nationwide sample. *Am J Clin Nutr* **68**, 266–274.

Kipnis V, Carroll RJ & Freedman LS (1999) Implications of a new dietary measurement error model for estimation of relative risk: application to four calibration studies. *Am J Epidemiol* **150**, 642–651.

Kipnis V, Subar AF, Midthune DN, Freedman LS, Ballard-Barbash R, Troiano RP, Bingham SA, Schoeller DA, Schatzkin & Carroll RJ (2003) Structure of dietary measurement error: results of the OPEN biomarker study. *Am J Epidemiol* **158**, 14–21.

Lilienthal Heitmann B & Lissner L (1995) Dietary underreporting by obese individuals – is it specific or non-specific? *Br Med J* **311**, 986–989.

Livingstone MBE & Black AE (2003) Markers of the validity of reported energy intake. *J Nutr* **133**, 895S–920S.

Löf M, Hannestad U & Forsum E (2003) Comparison of commonly used procedures, including the doubly-labelled water technique, in the estimation of total energy expenditure of women with special reference to the significance of body fatness. *Br J Cancer* **90**, 961–968.

Macdiarmid JI, Vail A, Cade JE & Blundell JE (1998) The sugar–fat relationship revisited: differences in consumption between men and women of varying BMI. *Int J Obes* **22**, 1053–1061.

Manjer J, Carlsson S, Elmståhl S, Gullberg B, Janzon L, Lindström M, Mattisson I & Bergland G (2001) The Malmö Diet and Cancer Study: representativity, cancer incidence and mortality in participants and non-participants. *Eur J Cancer Prev* **10**, 489–499.

Mattisson I, Wirfält E, Johansson U, Gullberg B, Olsson H & Berglund G (2004a) Intakes of plant foods and fat and risk of breast cancer in postmenopausal women – a prospective study in the Malmö Diet and Cancer cohort. *Br J Cancer* **94**, 122–127.

Mattisson I, Wirfält E, Wallström P, Gullberg B, Olsson H & Berglund G (2004b) High fat and alcohol intakes are risk factors of post-menopausal breast cancer: a prospective study in the Malmö Diet and Cancer Cohort. *Int J Cancer* **110**, 589–597.

National Bureau of Statistics (1989) *Occupations in Population and Housing Census 1985 (FoB 85) According to Nordic Standard Occupational Classification (Nordisk Yrkesklassificering, NYK) and Swedish Socioeconomic Classification (Socioekonomisk Indelning SEI)*. Reports on Statistical Co-ordination no. 1989:5. Stockholm: Statistics Sweden (in Swedish).

Price GM, Paul AA, Cole TJ & Wadsworth MEJ (1997) Characteristics of the low-energy reporters in a longitudinal national dietary survey. *Br J Nutr* **77**, 833–851.

Riboli E, Elmståhl S, Saracci R, Gullberg B & Lindgärde F (1997) The Malmö Food Study: validity of two dietary assessment methods for measuring nutrient intake. *Int J Epidemiol* **26**, S161–S173.

Richardsson MT, Leon AS, Jacobs DR Jr, Ainsworth BE & Serfass R (1994) Comprehensive evaluation of the Minnesota Leisure Time Physical Activity Questionnaire. *J Clin Epidemiol* **47**, 271–281.

Rosell MS, Hellénius M-L, de Faire UH & Johansson GK (2003) Associations between diet and the metabolic syndrome vary with the validity of dietary intake data. *Am J Clin Nutr* **78**, 84–90.

Salmerón J, Ascherio A, Rimm EB, Colditz GA, Spiegelman D, Jenkins DJ, Stampfer MJ, Wing AL & Willett WC (1997) Dietary fiber, glycemic load, and risk of NIDDM in men. *Diabetes Care* **20**, 545–550.

Seidell JC (1998) Dietary fat and obesity: an epidemiological perspective. *Am J Clin Nutr* **67**, Suppl., 546S–550S.

Stallone DD, Brunner E, Bingham SA & Marmot M (1997) Dietary assessment in Whitehall II: the influence of reporting bias on apparent socioeconomic variation in nutrient intakes. *Eur J Clin Nutr* **51**, 815–825.

Subar AF, Kipnis V, Troiano RP, *et al.* (2003) Using intake biomarker to evaluate the extent of dietary misreporting in a large sample of adults: the OPEN Study. *Am J Epidemiol* **158**, 1–13.

Taylor HL, Jacobs DR Jr., Schucker B, Knudsen J, Leon AS & Debacker G (1978) A questionnaire for the assessment of leisure time physical activities. *J Chronic Dis* **31**, 741–755.

Tooze JA, Subar AF, Thompson FE, Troiano RP, Schatzkin A & Kipnis V (2004) Psychosocial predictors of energy underreporting in a large doubly labeled water study. *Am J Clin Nutr* **79**, 795–804.

Warwick PM & Busby R (2004) Prediction of twenty-four-hour energy expenditure in a respiration chamber in smokers and non-smokers. *Eur J Clin Nutr* **47**, 600–603.

Willett WC, Howe GR & Kushi LH (1997) Adjustment for total energy intake in epidemiologic studies. *Am J Clin Nutr* **65**, Suppl., S1220–S1228.

Wirfält E, Mattisson I, Gullberg B & Berglund G (2000) Food patterns defined by cluster analysis and their utility as dietary exposure variables: a report from the Malmö Diet and Cancer Study. *Public Health Nutr* **3**, 159–173.

Wirfält E, Mattisson I, Johansson U, Gullberg B, Wallström P & Berglund G (2002) A methodological report from the Malmö Diet and Cancer study: development and evaluation of altered routines in dietary data processing. *Nutr J* **1**, 3, http://www.nutritionj.com.

World Health Organization Study Group (2000) *Obesity: Preventing and Managing the Global Epidemic. Report of a WHO Consultation.* WHO Technical Report Series no. 894. Geneva: WHO.

Young-Hwan J, Talmage DA & Role LW (2002) Nicotinic receptor-mediated effects on appetite and food intake. *Int J Neurobiol* **53**, 618–632.

Appendix 1

Intensity of leisure-time physical activities in the Malmö Diet and Cancer cohort

Physical activity	MET factor
Badminton, general	4·5
Table tennis	4·0
Soccer/handball	8·25
Golf	4·5
Jogging or running	7·5
Gymnastics	4·0
Orienteering	9·0
Swimming	6·0
Tennis	7·0
Cycling	4·0
Walking	3·5
Walking on stairs	5·5
Dancing (folk)	5·5
Dancing	3·0
Lawn mowing by hand	6·0
Digging the garden	5·0
Garden work	5·0

Activities under the open-ended option were recoded and entered as the most relevant activity, in terms of intensity, of the seventeen activities listed.

Appendix 2

Intensity levels of work

'How physically strenuous is your work? (version 1 of the questionnaire)

	MET factor
Sedentary (e.g. office work, watch making)	1·5
Somewhat strenuous (e.g. shop assistant, supervisor, teaching, light industrial work)	3·5
Moderately strenuous (e.g. postman, caretaker, heavy industrial work, building work)	4·5
Very strenuous (e.g. farm worker, heavy worker, fisherman)	6·5

What degree of physical activity is usually demanded in your work? (versions 2 and 3 of the questionnaire)

	MET factor
Very light (e.g. reading, office work, teaching)	1·5
Light (e.g. precision mechanics, feeding patients, portion medicine to patients, washing-up)	3·0
Medium heavy (e.g. washing, making beds, cleaning, child care, carpentry)	4·0
Heavy (e.g. lifting and turning patients, street-sweeper, heavy garden work, loading and unloading goods	5·0
Very heavy (e.g. concreting, shovelling earth or sand)	6·5