Title: The effect of under-reporting of energy intake on dietary patterns and on the associations between dietary patterns and self-reported chronic disease in women aged 50-69.

Authors: Marianne S. Markussen[1], Marit B. Veierød [1,2], Giske Ursin [1,3,4], Lene F. Andersen[1]

[1] University of Oslo, Institute of Basic Medical Sciences, Department of Nutrition
P.O. Box 1046 Blindern
0317 Oslo
Norway

[2] University of Oslo, Institute of Basic Medical Sciences Department of Biostatistics, Center for Biostatistics and Epidemiology
P.O. Box 1122 Blindern
0317 Oslo
Norway

[3] Cancer Registry of Norway
P.O. box 5313 Majorstuen
0304 Oslo
Norway

[4] Department of Preventive Medicine, University of Southern California
Soto Street Building
2001 N Soto Street
Los Angeles, CA 90032-3628
USA

Corresponding author:

Giske Ursin

E-Mail: giske.ursin@kreftregisteret.no

Address: See [3]

Phone: +47 22 92 87 01

Short title: Effect of under-reporting on dietary patterns and disease

Key words: dietary patterns; principal component analysis; misreporting of energy; low energy reporters; Goldberg cut-off

Abstract

The aim of this cross-sectional study was to investigate whether under-reporting of energy intake will affect derived dietary patterns and the association between the dietary patterns and self-reported chronic disease. Subjects were 6204 women aged 50-69 years participating in the Norwegian Breast Cancer Screening Program. Diet was assessed using a 253-item FFQ, and categorized into forty-nine food groups. We identified dietary patterns using principal component analysis. According to the revised Goldberg cut-off method, women with EI/BMR<1.10 was classified as low energy reporters. We examined the associations between dietary patterns and self-reported chronic diseases among all reporters and plausible reporters (low energy reporters excluded) by multivariable logistic regression. A total of 1133 (18%) women were low energy reporters. They reported higher body mass index, lower physical activity, lower alcohol intake, lower education and higher prevalence of some diseases compared to plausible reporters ($0.001 \leq P \leq 0.04$). We identified a “Prudent”, “Western” and “Continental” dietary pattern among both all and plausible reporters. The majority of the food groups identified in these patterns were consistently found among both these groups of women. The associations between dietary patterns and self-reported chronic disease became stronger, especially for the “Prudent” pattern, when analyses were restricted to plausible reporters. In summary we observed the strongest effect of under-reporting on these associations among the overweight/obese women. Under-reporting of energy intake attenuated the associations between dietary patterns and self-reported chronic disease, especially among the overweight/obese women. We suggest that it is important to consider the potential effect of measurement errors due to under-reporting on the effect estimate between dietary patterns and disease.

Introduction

The relationship between diet and chronic disease is complex. We consume foods and nutrients in different combinations and as part of meals. Thus, evaluating diet as a whole, based on dietary patterns is a complementary approach to the study of single nutrients or foods to understand the relationship between diet and disease [(1)]. When investigating associations between diet and disease the findings are in most cases based on self-reported dietary intake. Previous research has revealed extensive misreporting, especially under-reporting, of self-reported dietary intake [(2-5)]. The misreporting can be general under-reporting of food intake, or under- or over-reporting of certain food groups related to social desirability [(6-8)]. This latter misreporting may distort dietary patterns, and such a distortion could result in erroneous conclusions regarding the associations between dietary patterns and disease.

The doubly labeled water technique (DLW) has been looked upon as a gold standard in the evaluation of reported energy intake (EI). Unfortunately, the DLW method is technically challenging and extremely expensive and therefore not possible to implement in most studies. The more simple method developed by Goldberg et al. [(9)] and later revised by Black [(10)], has been proposed as an alternative to identify potential misreporters of energy intake. By using the level of discrepancy between the ratio of EI to the estimated basal metabolic rate (BMR) and the presumed physical activity level (PAL) of the population, the individuals can be classified as likely to be low energy-, plausible- or high energy reporters [(9, 10)].

A few studies have investigated the effect of under-reporting on empirically derived dietary patterns [(11-18)]. Four studies reported that the composition of food groups in the dietary patterns remained relatively unchanged after removal of low energy reporters [(11, 13-15)]. One study found that the number of dietary patterns differed between the plausible reporters and all reporters [(12)]. In all these studies cluster analysis was used to define dietary patterns. The distribution of low energy reporters across clusters was not uniform and whether the highest proportion of low energy reporters were found in the healthy or unhealthy clusters differed between the studies [(11, 12, 14-18)]. The cluster analysis assigns the study subjects to one of a number of discrete clusters or dietary patterns. When deriving dietary patterns by principal component analysis (PCA) an individual's diet is characterized using a continuous score for each of the derived patterns, thus this method has the advantage that it looks at more than one dimension of variation in the diet [(19)]. Recently, in a study among Swedish adults the researchers investigated the effect of excluding low energy reporters on dietary patterns derived by PCA [20] and found that the patterns were largely consistent. To the best of our knowledge that study is the only study that has investigated the effect of under-reporting on dietary patterns derived by PCA.

The aims of the present study were to investigate the effect of under-reporting of energy intake, by excluding low energy reporters from the study sample, on (a) the dietary patterns derived by PCA and (b) the association between the dietary patterns and self-reported chronic disease.

Methods

Study sample

The Norwegian Breast Cancer Screening Program is a governmentally funded national screening program administered by the Cancer Registry of Norway [21]. All Norwegian women aged 50-69 years are invited to a bilateral two-view mammogram biennially. The participation rate is 77% [22], with about 250,000 women invited per year. In 2006/2007, the Norwegian Breast Cancer Screening Program's invitation letter for mammographic screening included a question on willingness to complete a dietary questionnaire. A total of 67,527 women agreed to participate. In 2008, a consent form and a food frequency questionnaire (FFQ) were sent to a random sample of 10,000 of these women living all over Norway. A total of 6974 returned the FFQ, and 676 women were excluded because; the FFQs were not filled in (n=46); missing data on height and/or weight (n=158), age (n=5), smoking status (n=41), education (n=79), physical activity (n=104); height <125 cm (n=7), weight <30 kg or >170 kg (n=13); age not within the range 50-69 years (n=15); BMI <18.5 kg/m^2 or ≥40 kg/m^2 (n=98); or energy intake <2100 kJ/day or >15,000 kJ/day (n=204). This left us with a total sample of 6204 women.

This study was conducted according to the guidelines laid down in the Declaration of Helsinki and all procedures involving human subjects were approved by the regional ethics committee and the Norwegian Data Protection Authority. Written informed consent was obtained from all subjects.

Dietary assessment

The 16-page, 253-item FFQ was designed to measure the habitual food intake among Norwegian adults the preceding year. The questionnaire had an extra focus on fruit, vegetables and other antioxidant-rich foods and beverages, so the foods accounting for the variation in antioxidant intake in a population could be investigated [23]. The 253-item FFQ has been described in details earlier [24]. Shortly, it was based on a previously validated 180-item FFQ designed to measure the total energy intake in the Norwegian population [25], which later was expanded to a 270-item FFQ to cover the most antioxidant-rich foods and beverages in Norway [26]. The energy and food intake estimated from the 270-item FFQ has been validated [26, 27]. The energy intake was compared with independent measures of energy expenditure using the ActiReg® system (motion detection) [28], whereas 7-days weighed food records were used to study the relative validity of food and nutrient intake. The correlation coefficient between energy intake and energy expenditure was 0.54. [26]. The 253-item FFQ used in this study was revised from the original 270-item FFQ by removing 17 items that were seldom or never eaten (for example vegetables as curly kale, red cabbage, globe artichoke, sundried tomatoes and tofu; herbs and spices as cumin, turmeric, ginger powder, caraway, cloves, piri piri and sage). The questionnaire also collected information about dietary supplements, age, height, weight, smoking, physical activity, chronic diseases (present or previous) and medication. Daily intake of energy, nutrients and foods were computed using the food database AE-07 and KBS software system (KBS, version 4.9 2008) developed at the Department of Nutrition, University of Oslo, Norway. The food database AE-07 is based on the 2006 edition of the Norwegian food composition table (www.norwegianfoodcomp.no). Intakes from dietary supplements were included in the calculations.

The 253 food items were categorised into 49 food groups based on similarity in ingredients, nutrient profile or culinary usage (**Supplement Table 1**).

Disease assessment

In the FFQ, the participants were asked if they had currently or previously been diagnosed with one or more of the following diseases: asthma, joint inflammation, muscle or skeletal disorder, chronic gastrointestinal disease, chronic respiratory disease, depression or psychiatric disorder, stroke, heart attack or angina, hypertension and diabetes (type 1 or type

2). We defined six disease groups: total chronic disease (composed of all of the following disease groups), cardiovascular disease (stroke, heart attack, angina and hypertension), diabetes (type 1 and 2), chronic respiratory disease (asthma and chronic respiratory disease), cancer and joint/muscle/skeletal disorders (joint inflammation, muscle and skeletal disorders). A participant was identified to belong to a disease group if she had at least one of the diseases in the group.

Physical activity assessment

Physical activity was assessed using a modified version [29] of the physical activity questionnaire used in the California Teachers Study [30]. Subjects were asked to assess habitual weekly physical activity, and report all physical activity lasting at least 10 minutes per session. They were provided examples of light activities (defined as walking or cross-country skiing at a slow pace), moderate activities (defined as activities where some effort is required and which cause increased breathing, such as bicycling, swimming or cross-country skiing at a moderate pace, jogging at a slow pace, dancing) and strenuous activities (defined as activities that require hard effort and causes substantial increased breathing, such as aerobics, running, cross-country skiing or bicycling at a brisk pace). The subjects were asked to estimate their mean hours per week (none, <0.5, 0.5-1, 1.5-2, 2.5-3.5, 4-6, ≥7 hours) of participation at each level of activity. We created separate light, moderate and strenuous activity variables in minutes per week by summing up hours per week for each level of activity multiplied with 60.

We also calculated energy expenditure as the number of hours of each physical activity multiplied by its estimated metabolic cost [31] and expressed this variable in metabolic equivalent task (MET h/week).

Definition of low energy reporters

Low energy reporters were determined using the revised Goldberg cut-off method [9, 10]. This method is based on the principle that EI equals energy expenditure (EE) when weight is stable (equation (1): EI=EE). EE can also be expressed as multiples of BMR and physical activity level (PAL), and replacing EE in equation (1) with BMR x PAL gives equation (2): EI/BMR

= PAL. The idea by Goldberg and colleagues were that the ratio EI/BMR can be derived from a dietary assessment method and then be evaluated against an expected PAL for a population.

The revised Goldberg cut-offs [9, 10], used in the present study are based on estimated 95% confidence limits (cut-offs) for the plausible EI. The values of these cut-offs varies according to physical activity level (PAL), number of days of food recording and whether the evaluation of EI/BMR is at the individual or group level. Subjects are defined as plausible-, low energy- or high energy reporters from their ratio of EI/BMR according to whether this ratio are within, below or above the 95% confidence limits calculated, respectively.

We have used the lower 95% confidence limit published by Black [10] to identify low energy reporters, which is based on a PAL of 1.55, and infinity number of days of food recording (habitual intake measured by a FFQ) at the individual level. The value of this cut-off is 1.10, therefore, all women with EI/BMR<1.10 was classified as low energy reporters in this study.

In the present study BMR is calculated from the following equations [32]:

BMR women 31-60 years: $0.0433\ W + 2.57\ H – 1.180$

and

BMR women 61-70 years: $0.0342\ W + 2.10\ H – 0.0486$,

Statistical methods

We divided the study sample into all and plausible reporters, and each of these subsamples was stratified by BMI; 18.5 kg/m^2≤BMI<25 kg/m^2 (normal weight) and 25 kg/m^2≤BMI<40 kg/m^2 (overweight/obese).

Differences between under- and plausible reporters, and between normal weight and overweight/obese were studied by two-sample t-test for continuous variables and chi-square test for categorical variables. Physical activity and alcohol intake data was $\log_e$ transformed in these analyses.

Principal component analysis (PCA) was used to derive dietary patterns based on the correlation matrix of the 49 food groups (g/day). Prior to extracting components, the suitability for using PCA was assessed by the Kaiser-Meyer-Olkin measure of sampling adequacy and the Bartlett's test of sphericity which tests whether our correlation matrix is significantly different from an identity matrix [33]. The Kaiser-Meyer-Olkin value was 0.76 for both all and plausible reporters, which is above the suggested minimum of 0.50 [34], and Bartlett's test of sphericity was statistically significant (P<0.001), supporting the suitability of the data for PCA. The input variables were standardized by using the correlation matrix of the 49 food group variables in the PCA, and not the covariance matrix. To determine the number of meaningful components to retain, we considered the eigenvalue-one criterion, the scree test, the proportion of variance accounted for and the interpretability of the patterns [35]. For interpretation purposes, varimax rotation was performed on the retained components. We considered food groups with a factor loading ≥0.3 (absolute value) to load on that component. We interpreted the retained components as dietary patterns and labelled them according to the more or less healthy combinations of food groups. Finally, each woman's score was calculated for each of the retained components.

The association between dietary pattern scores and the prevalence of diseases among all and plausible reporters were estimated using a logistic regression model. The dietary pattern scores were categorised into tertiles and we estimated the adjusted odds ratios (ORs) and 95% confidence intervals (CIs) for each tertile compared with the lowest tertile of each dietary pattern. We analysed trends across tertiles of dietary pattern scores by treating the variable as a continuous variable in the regression analysis. Adjustment for age (continuous), education (categorical), smoking (yes/no), physical activity (continuous) and energy intake (continuous) was made. We tested for interaction between BMI (two categories) and dietary pattern scores.

All tests were two sided and P<0.05 was considered statistically significant. The analyses were conducted using SPSS version 20.0 (IBM Corp., Somers, New York, USA).

Results

A total of 1133 (18.3%) of the 6204 women were defined as low energy reporters (**Table 1**). Low energy reporters had significantly lower energy intake, higher BMI, lower physical activity, lower alcohol intake and lower education than plausible reporters (P≤0.02). Also, there was an indication of a higher proportion of smokers among low energy reporters than among plausible reporters (P=0.09). The prevalence of self-reported total chronic disease, cardiovascular disease, diabetes and joint/muscle/skeletal disorders was significantly higher

($P \leq 0.04$) among low energy reporters compared to plausible reporters. Overweight/obesity was more common in low energy reporters than plausible reporters (62.8% and 45.8%, respectively).

Among all reporters the energy intake was significantly higher for the normal weight group than for the overweight/obese group ($P=0.02$) (**Table 2**). By removal of the low energy reporters this changed to the opposite ($P=0.001$). In both all reporters and plausible reporters the overweight/obese group was slightly older, had lower physical activity, lower alcohol intake, were less likely to smoke, were less educated and had a higher prevalence of chronic diseases than the normal weight group ($P<.001$).

We identified three major dietary patterns for both all reporters and plausible reporters, all with eigenvalues ≥2.0. The point at which the slope of the graph in the scree plot showed a change, and the interpretation of the components, justified retaining three components. **Table 3** presents the three dietary patterns for all and plausible reporters, with food groups having factor loadings with absolute values ≥0.30 in bold. The three dietary patterns accounted for 17.4% and 16.7% of the total variance among all and plausible reporters, respectively. Among all reporters the dietary pattern labelled "Prudent" was characterised by high positive loadings for vegetables, fish as dinner, fruits, herbs and spices, berries, nuts and seeds, legumes, meat dishes, salad dressings, poultry, vegetarian food, soup, and tea. Although the "Prudent" pattern derived for the plausible reporters was substantially similar to that of all reporters, differences were noted for three food groups: Vegetarian food, tea and salad dressings which had no longer factor loadings ≥30. Furthermore, the "Prudent" pattern explained the highest amount of variance in dietary intake among all reporters, whereas among plausible the "Western" pattern explained the highest amount of variance. Among all reporters the "Western" dietary pattern was characterised by high loadings for potatoes, sauce, refined grains, processed meat, cakes and desserts, margarine, sweet spreads, red meat and game, and high negative loadings for wine and herbs and spices. For plausible reporters a similar "Western" pattern was found, but this pattern also showed a high negative loading for vegetarian food. The "Western" pattern had the highest total variance explained among the plausible reporters. The third pattern was labelled "Continental" and among all reporters it was characterised by high loadings for tomato sauce, pasta, processed meat, fat-rich potatoes, salty snacks, pizza, salad dressings, rice, poultry, mustard, sweets and wine. We found a similar "Continental" pattern among plausible reporters, with soy sauce among the high loaded food groups (0.31) and wine with slightly less loading (0.29).

We found significant interaction between dietary pattern score and BMI in the analyses of the associations between dietary pattern scores and self-reported chronic diseases in all and plausible reporters ($P \leq 0.005$), thus the results are presented stratified by BMI.

Table 4 presents the adjusted ORs of self-reported chronic disease by tertiles of the dietary pattern scores among all and plausible reporters with normal weight. Among plausible reporters the “Prudent” pattern was significantly positively associated with self-reported total chronic disease [odds ratio (OR) for highest compared to lowest tertile: 1.43; 95% CI: 1.14, 1.80; P_{trend}=0.002]. Among all reporters the effect estimates were attenuated and no longer significant, however the trend was still significant [OR for highest compared to lowest tertile: 1.24; 95% CI: 1.00, 1.55; P_{trend}=0.05]. The “Prudent” pattern was significantly positively associated with joint/muscle/skeletal disorder among both plausible reporters and all reporters, but also here the effect estimates were attenuated among all reporters compared to plausible reporters [OR for highest compared to lowest tertile for; plausible reporters: 1.69; 95% CI: 1.31, 2.18; P_{trend}<0.001; all reporters: 1.44; 95% CI: 1.13, 1.85; P_{trend}= 0.003]. The “Continental” pattern was inversely associated with joint/muscle/skeletal disorder among both plausible and all reporters [OR for highest compared to lowest tertile for; plausible reporters: 0.77; 95% CI: 0.62, 0.97; P_{trend}=0.04; all reporters: 0.77; 95% CI: 0.62, 0.95; P_{trend}=0.02], but the trend was weaker among plausible reporters.

Table 5 presents the adjusted ORs of self-reported chronic disease by tertiles of the dietary pattern scores among overweight/obese all and plausible reporters. The “Prudent” pattern was significantly positively associated with total chronic disease, and the OR was only slightly higher in highest related to lowest tertile for plausible reporters compared to all reporters [OR for highest compared to lowest tertile for; plausible reporters: 1.46; 95% CI: 1.14, 1.87; P_{trend}=0.003; all reporters: 1.45; 95% CI: 1.14, 1.84; P_{trend}=0.003]. When we looked at each disease separately, we found that there was a significant positive association between the “Prudent” pattern and cardiovascular disease among both plausible reporters and all reporters [OR for highest compared to lowest tertile for plausible reporters: 1.63; 95% CI: 1.19, 2.23; P_{trend}=0.002; all reporters: 1.69; 95% CI: 1.23, 2.27; P_{trend}=0.001]. The “Prudent” pattern was significantly positively associated with diabetes among both plausible and all reporters [OR for highest compared to lowest tertile for; plausible reporters: 3.82; 95% CI: 1.95, 7.51; P_{trend}<0.001; all reporters: 2.80; 95% CI: 1.62, 4.87; P_{trend}<0.001], and with chronic respiratory disease among plausible reporters [OR for highest compared to lowest tertile: 1.62; 95% CI: 1.09, 2.40; P_{trend}=0.02]. The “Western” pattern was significantly positively associated with cancer among plausible reporters [OR for highest compared to lowest tertile: 1.68; 95% CI: 1.02, 2.77; P_{trend}=0.03]. The effect estimates observed between the “Prudent” as well as the “Western” pattern and diabetes, chronic respiratory disease and cancer among plausible reporters were all attenuated among all reporters. Finally, the “Prudent” pattern was also significantly positively associated to joint/muscle/skeletal disorder among both plausible and all reporters, however the effect estimate and P_{trend} was weaker among plausible reporters [OR for highest compared to lowest tertile for; plausible reporters: 1.33; 95% CI: 1.01, 1.75; P_{trend}=0.04; all reporters: 1.44; 95% CI: 1.11, 1.87; P_{trend}=0.007].

The highest effects of under-reporting on the associations between dietary patterns and self-reported chronic diseases were observed among the overweight/obese women.

Supplemental Table 2 shows the effect of including the covariates one by one in the logistic regression model of the relationship between the tertiles of dietary pattern score and self-reported total chronic disease among plausible reporters. We observed no significant associations between self-reported total chronic disease and the “Western” and the “Continental” pattern. Among the normal-weight women the association between the “Prudent” pattern and self-reported total chronic disease changed most, while the ORs for the association among the overweight/obese remained unchanged when including the covariates in the model.

Discussion

We identified almost one fifth of the women to be low energy reporters based on the revised Goldberg cut-off method [9, 10]. The majority of the food groups identified in the “Prudent”, “Western” and “Continental” patterns were consistently found for both all and plausible reporters, differing only with a few food groups. Due to statistically significant interaction between dietary pattern score and BMI we stratified the women in two groups, normal-weight and overweight/obese. We observed more statistical significant associations between dietary patterns and self-reported chronic diseases among the overweight/obese than the normal weight women. More importantly, the associations between dietary patterns and self-reported chronic diseases became stronger when analyses were restricted to plausible reporters. Specifically the associations between the “Prudent” pattern and self-reported chronic diseases strengthened.

Studies using the DLW method have clearly shown that all dietary assessment methods tend to underestimate energy intake to various degrees [36-38]. Previous studies have reported prevalence of low energy reporting ranging from 10 to 60% depending on the dietary assessment method, the reference method used to identify low energy reporters and the characteristics of the study population [4, 11, 37, 39-51]. In the revised Goldberg cut-off equations [10], the individual's physical activity are taken into account. To increase sensitivity Black [10] recommended collecting more information about home or occupational and leisure time physical activity, to be able to assign subjects to low, medium and high activity categories. Three different cut-offs can then be calculated for the subjects belonging to the different activity categories. This would probably have resulted in a higher prevalence of low energy reporters in our study sample. Unfortunately, the physical activity questionnaires used in the present study did not give enough information about the individuals' total amount of physical activity. Therefore, we used a PAL of 1.55, which is the value defined by FAO/WHO/United Nations University representing a sedentary level of energy expenditure [52], in order not to over-estimate the extent of under-reporting. Though, it could be criticised to be a too conservative PAL value for this population, and misclassifications of more active participants could exist. We found a prevalence of 18.3% low energy reporters in our study sample, which was somewhat lower than in comparable studies reporting a prevalence of 25 to 38% [37, 53-57]. In these studies less detailed FFQ's (70-180 food items) were used to collect dietary data, and higher cut-off values were used for identifying low energy reporters (1.14-1.35) compared to our study. In addition, we had deficient information of total physical activity in our study. It is important to take into account that the confidence limits calculated by Goldberg and Black are wide, and only extreme degrees of misreporting can be identified [10].

The low energy reporters in this study reported higher BMI, lower physical activity, lower alcohol intake and lower education than the plausible reporters (Table 1). This is in line with previous studies investigating characteristics of low energy reporters [13, 37, 58].

We have previously discussed the dietary patterns derived in this study in detail [59]. In the current analyses, we wanted to investigate if the measurement errors introduced by under-reporting distorted the food groups defining the dietary patterns. We found three major dietary patterns among both all reporters and plausible reporters that were not vastly different from each other, differing only with a few food groups in each pattern. Other studies have also identified relatively similar patterns after removal of low energy reporters from the analysis compared to the total sample [13-15]. Interestingly, the dietary pattern explaining the highest total variance differed between all and plausible reporters, with the "Prudent" pattern explaining the highest amount of variance in dietary intake among all reporters, and the "Western" pattern explaining the highest amount of variance among plausible reporters. This may be related to the fact that low energy reporters tend to over-report foods perceived as healthy or/and under-report foods perceived as unhealthy [8, 42, 60, 61]. The identification of the first principal component as a prudent dietary pattern among all reporters is comparable with other studies investigating dietary patterns derived by PCA [19, 62-65]. Some, but not all [18], studies investigating the association between dietary patterns derived by cluster analysis and under-reporting of energy intake observed more severe under-reporting among subjects in healthy dietary pattern clusters [16, 66, 67].

The implications of under-reporting might be distortion of the associations between diet and disease. Most studies in nutritional epidemiology have excluded subjects with very high or very low energy intake frequently using the cut-offs <2100 kJ/day and >15,000 kJ/day, however this does not account for all the misreporting [68-73]. As found in this study, the associations between dietary patterns and self-reported disease differed between all reporters (excluding those with implausible energy intake <2100 kJ/day and >15,000 kJ/day), and plausible reporters (additional exclusion of low energy reporters as defined by the revised Goldberg cut-off method [10]). Results showed that the associations between dietary patterns and self-reported chronic diseases generally became stronger when analyses were restricted to plausible reporters, and especially among the overweight/obese. Particularly the associations between the “Prudent” pattern and self-reported chronic diseases strengthened. The positive relationship between the “Prudent” pattern and several of the diseases indicated that the participants tried to eat healthy in order to reduce either the symptoms of their condition, or reduce the likelihood of possible detrimental consequences. Positive relationship between a healthy dietary pattern and disease has also been reported in a Swedish study, where the highest prevalence of previously known health problems was observed in the healthy “Fruit & vegetables” cluster among women [66].

A Swedish study investigated the effect of under-reporting on the association between risk of breast cancer and alcohol intake [74]. The researchers reported an increased risk of breast cancer with high alcohol intakes, and the risk estimates were strengthened among the plausible reporters compared to all reporters. A study in the US [75] investigated the use of calibrated energy intake to account for under-reporting and the effect on the association between risk of breast, colon, endometrial and kidney cancer. They produced the calibrated consumption estimates based on calibration equations developed in a substudy among 544 women where DLW was used to estimate total energy expenditure and urinary nitrogen was used as recovery biomarker for protein [76]. The researchers found calibrated energy consumption to be positively associated with the risk of breast, colon, endometrial, and kidney cancer, whereas uncalibrated energy was not.In a few studies, the investigators have adjusted for under-reporting of energy intake in their analyses in order to avoid biased conclusions [77-79].

The extensive information on diet, lifestyle and self-reported chronic diseases and the large study sample from different parts of the country are important strengths of the present study. However, there are some limitations. Firstly, it might be that the women responding to the FFQ were healthier and/or more health conscious than those not responding. Secondly, since the FFQ had an extra focus on fruit and vegetables, these food items may have been overestimated. Thirdly, when deriving dietary patterns by PCA many subjective decisions are made that also can impact the number and type of dietary patterns [1, 80-82]. Fourthly, the sensitivity of the revised Goldberg cut-off method increases if it is possible to assign participants to low, medium and high physical activity categories based on the total amount of physical activity. Unfortunately, only recreational light, moderate and vigorous physical activity was assessed, and not occupational or household physical activities, which are important contributors to total energy expenditure. If it had been possible to calculate three different cut-offs based on three different PALs, we might have found a higher prevalence of low energy reporters in study [10]. However, the questionnaire used in the present work was designed for a study on diet and breast cancer, and the questions were designed based on that recreational physical activity has been found to have the strongest association with risk of breast cancer [83].

In conclusion, in this large sample of women aged 50-69 we identified three dietary patterns: “Prudent”, “Western” and “Continental” for both all and plausible reporters. The food group composition of the dietary patterns were quite similar for both all and plausible reporters, however, the pattern contributing most to the explanation of variances in the food groups was the “Prudent” among all reporters and the “Western” among plausible reporters. We also found that under-reporting of energy intake attenuated the associations between dietary patterns and self-reported chronic diseases, especially among overweight/obese women. . We suggest that it is important to consider the potential effect of measurement errors due to under-reporting on the effect estimate between dietary patterns and disease.

Aknowledgements

Financial support

This project was supported by the Research Council of Norway (G.U., grant number 196999). The Research Council of Norway had no role in the design, analysis or writing of this article.

Conflict of interest

None of the authors have any financial or other interests concerning the outcomes of the investigation or any conflicts of interest to declare.

Authorship

MSM carried out the calculation of the daily intake of energy, nutrients and foods, the statistical analyses and drafted the manuscript. MBV contributed to the statistical analysis and the interpretation of the data and the revisions of the manuscript. GU designed the study, obtained the funding, planned and executed the data collection, participated in discussion of results and development of the manuscript. LFA contributed to the interpretation of the data and the revisions of the manuscript. All authors read and approved the final manuscript.

References

1. Newby PK, Tucker KL (2004) Empirically derived eating patterns using factor or cluster analysis: a review. *Nutr Rev* **62**, 5, 177-203.
2. Schoeller DA (1995) Limitations in the assessment of dietary energy intake by self-report. *Metabolism* **44**, 2 Suppl 2, 18-22.
3. Schoeller DA (1990) How accurate is self-reported dietary energy intake? *Nutr Rev* **48**, 10, 373-379.
4. Subar AF, Kipnis V, Troiano RP*, et al.* (2003) Using intake biomarkers to evaluate the extent of dietary misreporting in a large sample of adults: the OPEN study. *Am J Epidemiol* **158**, 1, 1-13.
5. Jonnalagadda SS, Mitchell DC, Smiciklas-Wright H*, et al.* (2000) Accuracy of energy intake data estimated by a multiple-pass, 24-hour dietary recall technique. *J Am Diet Assoc* **100**, 3, 303-308; quiz 309-311.
6. Connor Gorber S, Tremblay MS (2010) The bias in self-reported obesity from 1976 to 2005: a Canada-US comparison. *Obesity (Silver Spring)* **18**, 2, 354-361.
7. Hebert JR, Hurley TG, Peterson KE*, et al.* (2008) Social desirability trait influences on self-reported dietary measures among diverse participants in a multicenter multiple risk factor trial. *J Nutr* **138**, 1, 226s-234s.
8. Scagliusi FB, Polacow VO, Artioli GG*, et al.* (2003) Selective underreporting of energy intake in women: magnitude, determinants, and effect of training. *J Am Diet Assoc* **103**, 10, 1306-1313.
9. Goldberg GR, Black AE, Jebb SA*, et al.* (1991) Critical evaluation of energy intake data using fundamental principles of energy physiology: 1. Derivation of cut-off limits to identify under-recording. *Eur J Clin Nutr* **45**, 12, 569-581.
10. Black AE (2000) 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 Relat Metab Disord* **24**, 9, 1119-1130.
11. Funtikova AN, Gomez SF, Fito M*, et al.* (2015) Effect of Energy Under-Reporting on Secular Trends of Dietary Patterns in a Mediterranean Population. *PLoS One* **10**, 5, e0127647.
12. Shaneshin M, Jessri M, Rashidkhani B (2014) Validity of energy intake reports in relation to dietary patterns. *J Health Popul Nutr* **32**, 1, 36-45.
13. Bailey RL, Mitchell DC, Miller C*, et al.* (2007) Assessing the effect of underreporting energy intake on dietary patterns and weight status. *J Am Diet Assoc* **107**, 1, 64-71.
14. Winkvist A, Hornell A, Hallmans G*, et al.* (2009) More distinct food intake patterns among women than men in northern Sweden: a population-based survey. *Nutr J* **8**, 12.
15. Martikainen P, Brunner E, Marmot M (2003) Socioeconomic differences in dietary patterns among middle-aged men and women. *Soc Sci Med* **56**, 7, 1397-1410.
16. Scagliusi FB, Ferriolli E, Pfrimer K*, et al.* (2008) Under-reporting of energy intake is more prevalent in a healthy dietary pattern cluster. *Br J Nutr* **100**, 5, 1060-1068.
17. Holmback I, Ericson U, Gullberg B*, et al.* (2009) Five meal patterns are differently associated with nutrient intakes, lifestyle factors and energy misreporting in a sub-sample of the Malmo Diet and Cancer cohort. *Food Nutr Res* **53**.
18. Pryer JA, Nichols R, Elliott P*, et al.* (2001) Dietary patterns among a national random sample of British adults. *J Epidemiol Community Health* **55**, 1, 29-37.
19. Crozier SR, Robinson SM, Borland SE*, et al.* (2006) Dietary patterns in the Southampton Women's Survey. *Eur J Clin Nutr* **60**, 12, 1391-1399.
20. Ax E, Warensjo Lemming E, Becker W*, et al.* (2015) Dietary patterns in Swedish adults; results from a national dietary survey. *Br J Nutr*, 1-10.
21. Hofvind SS, Wang H, Thoresen S (2003) The Norwegian Breast Cancer Screening Program: re-attendance related to the women's experiences, intentions and previous screening result. *Cancer Causes Control* **14**, 4, 391-398.

22. Hofvind S, Geller B, Vacek PM*, et al.* (2007) Using the European guidelines to evaluate the Norwegian Breast Cancer Screening Program. *Eur J Epidemiol* **22**, 7, 447-455.
23. Qureshi SA, Lund AC, Veierod MB*, et al.* (2014) Food items contributing most to variation in antioxidant intake; a cross-sectional study among Norwegian women. *BMC Public Health* **14**, 45.
24. Markussen MS, Veierod MB, Sakhi AK*, et al.* (2015) Evaluation of dietary patterns among Norwegian postmenopausal women using plasma carotenoids as biomarkers. *Br J Nutr*, 1-11.
25. Andersen LF, Solvoll K, Johansson LR*, et al.* (1999) Evaluation of a food frequency questionnaire with weighed records, fatty acids, and alpha-tocopherol in adipose tissue and serum. *Am J Epidemiol* **150**, 1, 75-87.
26. Carlsen MH, Lillegaard IT, Karlsen A*, et al.* (2010) Evaluation of energy and dietary intake estimates from a food frequency questionnaire using independent energy expenditure measurement and weighed food records. *Nutr J* **9**, 37.
27. Carlsen MH, Karlsen A, Lillegaard IT*, et al.* (2011) Relative validity of fruit and vegetable intake estimated from an FFQ, using carotenoid and flavonoid biomarkers and the method of triads. *Br J Nutr* **105**, 10, 1530-1538.
28. Berntsen S, Hageberg R, Aandstad A*, et al.* (2010) Validity of physical activity monitors in adults participating in free-living activities. *Br J Sports Med* **44**, 9, 657-664.
29. Qureshi SA, Ellingjord-Dale M, Hofvind S*, et al.* (2012) Physical activity and mammographic density in a cohort of postmenopausal Norwegian women; a cross-sectional study. *Springerplus* **1**, 1, 75.
30. Dallal CM, Sullivan-Halley J, Ross RK*, et al.* (2007) Long-term recreational physical activity and risk of invasive and in situ breast cancer: the California teachers study. *Arch Intern Med* **167**, 4, 408-415.
31. Ainsworth BE, Haskell WL, Whitt MC*, et al.* (2000) Compendium of physical activities: an update of activity codes and MET intensities. *Med Sci Sports Exerc* **32**, 9 Suppl, S498-504.
32. Henry CJ (2005) Basal metabolic rate studies in humans: measurement and development of new equations. *Public Health Nutr* **8**, 7a, 1133-1152.
33. Tabachnick BG FL (2007) Using Multivariate Statistics. 5th ed. p: Pearson Education Publications.
34. Hutcheson G, Sofroniou N (1999) The multivariate social scientist. p. London: Sage.
35. Michels KB, Schulze MB (2005) Can dietary patterns help us detect diet-disease associations? *Nutr Res Rev* **18**, 2, 241-248.
36. Hill RJ, Davies PS (2001) The validity of self-reported energy intake as determined using the doubly labelled water technique. *Br J Nutr* **85**, 4, 415-430.
37. Livingstone MB, Black AE (2003) Markers of the validity of reported energy intake. *J Nutr* **133 Suppl 3**, 895s-920s.
38. Schoeller DA, Schoeller DA (2002) Validation of habitual energy intake. *Public Health Nutr* **5**, 6a, 883-888.
39. Black AE, Goldberg GR, Jebb SA*, et al.* (1991) Critical evaluation of energy intake data using fundamental principles of energy physiology: 2. Evaluating the results of published surveys. *Eur J Clin Nutr* **45**, 12, 583-599.
40. Salle A, Ryan M, Ritz P (2006) Underreporting of food intake in obese diabetic and nondiabetic patients. *Diabetes Care* **29**, 12, 2726-2727.
41. Yannakoulia M, Panagiotakos DB, Pitsavos C*, et al.* (2007) Low energy reporting related to lifestyle, clinical, and psychosocial factors in a randomly selected population sample of Greek adults: the ATTICA Study. *J Am Coll Nutr* **26**, 4, 327-333.
42. Rasmussen LB, Matthiessen J, Biltoft-Jensen A*, et al.* (2007) Characteristics of misreporters of dietary intake and physical activity. *Public Health Nutr* **10**, 3, 230-237.
43. Poslusna K, Ruprich J, de Vries JH*, et al.* (2009) Misreporting of energy and micronutrient intake estimated by food records and 24 hour recalls, control and adjustment methods in practice. *Br J Nutr* **101 Suppl 2**, S73-85.

44. Mendez MA, Popkin BM, Buckland G*, et al.* (2011) Alternative methods of accounting for underreporting and overreporting when measuring dietary intake-obesity relations. *Am J Epidemiol* **173**, 4, 448-458.
45. Ferrari P, Slimani N, Ciampi A*, et al.* (2002) Evaluation of under- and overreporting of energy intake in the 24-hour diet recalls in the European Prospective Investigation into Cancer and Nutrition (EPIC). *Public Health Nutr* **5**, 6B, 1329-1345.
46. Rennie KL, Coward A, Jebb SA (2007) Estimating under-reporting of energy intake in dietary surveys using an individualised method. *Br J Nutr* **97**, 6, 1169-1176.
47. Becker W, Foley S, Shelley E*, et al.* (1999) Energy under-reporting in Swedish and Irish dietary surveys: implications for food-based dietary guidelines. *Br J Nutr* **81 Suppl 2**, S127-131.
48. Murakami K, Sasaki S, Okubo H (2012) Characteristics of under- and over-reporters of energy intake among young Japanese women. *J Nutr Sci Vitaminol (Tokyo)* **58**, 4, 253-262.
49. Freisling H, van Bakel MM, Biessy C*, et al.* (2012) Dietary reporting errors on 24 h recalls and dietary questionnaires are associated with BMI across six European countries as evaluated with recovery biomarkers for protein and potassium intake. *Br J Nutr* **107**, 6, 910-920.
50. Koebnick C, Wagner K, Thielecke F*, et al.* (2005) An easy-to-use semiquantitative food record validated for energy intake by using doubly labelled water technique. *Eur J Clin Nutr* **59**, 9, 989-995.
51. Luhrmann PM, Herbert BM, Neuhauser-Berthold M (2001) Underreporting of energy intake in an elderly German population. *Nutrition* **17**, 11-12, 912-916.
52. FAO/WHO/UNU (2001). Human energy requirements.
53. Lutomski JE, van den Broeck J, Harrington J*, et al.* (2011) Sociodemographic, lifestyle, mental health and dietary factors associated with direction of misreporting of energy intake. *Public Health Nutr* **14**, 3, 532-541.
54. Samaras K, Kelly PJ, Campbell LV (1999) Dietary underreporting is prevalent in middle-aged British women and is not related to adiposity (percentage body fat). *Int J Obes Relat Metab Disord* **23**, 8, 881-888.
55. Johansson L, Solvoll K, Bjorneboe GE*, et al.* (1998) Under- and overreporting of energy intake related to weight status and lifestyle in a nationwide sample. *Am J Clin Nutr* **68**, 2, 266-274.
56. Pomerleau J, Ostbye T, Bright-See E (1999) Potential underreporting of energy intake in the Ontario Health Survey and its relationship with nutrient and food intakes. *Eur J Epidemiol* **15**, 6, 553-557.
57. Smith WT, Webb KL, Heywood PF (1994) The implications of underreporting in dietary studies. *Aust J Public Health* **18**, 3, 311-314.
58. Park HA, Lee JS, Kuller LH (2007) Underreporting of dietary intake by body mass index in premenopausal women participating in the Healthy Women Study. *Nutr Res Pract* **1**, 3, 231-236.
59. Markussen MS, Veierod MB, Kristiansen AL*, et al.* (2015) Dietary patterns of postmenopausal women and associations with nutrient intake, sociodemographic factors and key risk factors for noncommunicable diseases. *Submitted*.
60. Lafay L, Mennen L, Basdevant A*, et al.* (2000) Does energy intake underreporting involve all kinds of food or only specific food items? Results from the Fleurbaix Laventie Ville Sante (FLVS) study. *Int J Obes Relat Metab Disord* **24**, 11, 1500-1506.
61. Krebs-Smith SM, Graubard BI, Kahle LL*, et al.* (2000) Low energy reporters vs others: a comparison of reported food intakes. *Eur J Clin Nutr* **54**, 4, 281-287.
62. Williams DE, Prevost AT, Whichelow MJ*, et al.* (2000) A cross-sectional study of dietary patterns with glucose intolerance and other features of the metabolic syndrome. *Br J Nutr* **83**, 3, 257-266.
63. Link LB, Canchola AJ, Bernstein L*, et al.* (2013) Dietary patterns and breast cancer risk in the California Teachers Study cohort. *Am J Clin Nutr* **98**, 6, 1524-1532.

64. Chocano-Bedoya PO, O'Reilly EJ, Lucas M*, et al.* (2013) Prospective study on long-term dietary patterns and incident depression in middle-aged and older women. *Am J Clin Nutr* **98**, 3, 813-820.
65. Herber-Gast GC, Mishra GD (2013) Fruit, Mediterranean-style, and high-fat and -sugar diets are associated with the risk of night sweats and hot flushes in midlife: results from a prospective cohort study. *Am J Clin Nutr* **97**, 5, 1092-1099.
66. Wirfalt E, Mattisson I, Gullberg B*, et al.* (2000) Food patterns defined by cluster analysis and their utility as dietary exposure variables: a report from the Malmo Diet and Cancer Study. *Public Health Nutr* **3**, 2, 159-173.
67. Hornell A, Winkvist A, Hallmans G*, et al.* (2010) Mis-reporting, previous health status and health status of family may seriously bias the association between food patterns and disease. *Nutr J* **9**, 48.
68. Anderson AL, Harris TB, Houston DK*, et al.* (2010) Relationships of dietary patterns with body composition in older adults differ by gender and PPAR-gamma Pro12Ala genotype. *Eur J Nutr* **49**, 7, 385-394.
69. Flores M, Macias N, Rivera M*, et al.* (2010) Dietary patterns in Mexican adults are associated with risk of being overweight or obese. *J Nutr* **140**, 10, 1869-1873.
70. Villegas R, Yang G, Gao YT*, et al.* (2010) Dietary patterns are associated with lower incidence of type 2 diabetes in middle-aged women: the Shanghai Women's Health Study. *Int J Epidemiol* **39**, 3, 889-899.
71. Sanchez-Villegas A, Delgado-Rodriguez M, Martinez-Gonzalez MA*, et al.* (2003) Gender, age, socio-demographic and lifestyle factors associated with major dietary patterns in the Spanish Project SUN (Seguimiento Universidad de Navarra). *Eur J Clin Nutr* **57**, 2, 285-292.
72. Song Y, Joung H (2012) A traditional Korean dietary pattern and metabolic syndrome abnormalities. *Nutr Metab Cardiovasc Dis* **22**, 5, 456-462.
73. Tucker KL, Chen H, Hannan MT*, et al.* (2002) Bone mineral density and dietary patterns in older adults: the Framingham Osteoporosis Study. *Am J Clin Nutr* **76**, 1, 245-252.
74. Mattisson I, Wirfalt E, Aronsson CA*, et al.* (2005) Misreporting of energy: prevalence, characteristics of misreporters and influence on observed risk estimates in the Malmo Diet and Cancer cohort. *Br J Nutr* **94**, 5, 832-842.
75. Prentice RL, Shaw PA, Bingham SA*, et al.* (2009) Biomarker-calibrated energy and protein consumption and increased cancer risk among postmenopausal women. *Am J Epidemiol* **169**, 8, 977-989.
76. Neuhouser ML, Tinker L, Shaw PA*, et al.* (2008) Use of recovery biomarkers to calibrate nutrient consumption self-reports in the Women's Health Initiative. *Am J Epidemiol* **167**, 10, 1247-1259.
77. McNaughton SA, Mishra GD, Brunner EJ (2009) Food patterns associated with blood lipids are predictive of coronary heart disease: the Whitehall II study. *Br J Nutr* **102**, 4, 619-624.
78. McNaughton SA, Mishra GD, Brunner EJ (2008) Dietary patterns, insulin resistance, and incidence of type 2 diabetes in the Whitehall II Study. *Diabetes Care* **31**, 7, 1343-1348.
79. Brunner EJ, Mosdol A, Witte DR*, et al.* (2008) Dietary patterns and 15-y risks of major coronary events, diabetes, and mortality. *Am J Clin Nutr* **87**, 5, 1414-1421.
80. Hu FB (2002) Dietary pattern analysis: a new direction in nutritional epidemiology. *Curr Opin Lipidol* **13**, 1, 3-9.
81. Northstone K, Ness AR, Emmett PM*, et al.* (2008) Adjusting for energy intake in dietary pattern investigations using principal components analysis. *Eur J Clin Nutr* **62**, 7, 931-938.
82. Smith AD, Emmett PM, Newby PK*, et al.* (2013) Dietary patterns obtained through principal components analysis: the effect of input variable quantification. *Br J Nutr* **109**, 10, 1881-1891.
83. Lynch BM, Neilson HK, Friedenreich CM (2011) Physical activity and breast cancer prevention. *Recent Results Cancer Res* **186**, 13-42.

Table 1. Selected characteristics and prevalence of disease of all-, under- and plausible reporters
(Number and percentage; means and SD; medians and 25th and 75th percentiles (P))

	All reporters		Low energy reporters		Plausible reporters		
Characteristics	Mean or median	SD or P25, P75	Mean or median	SD or P25, P75	Mean or median	SD or P25, P75	*P* *
n	6204		1133		5071		
Age (years)	57.9	4.5	57.6	4.5	57.9	4.5	0.08
Energy intake (kJ/day)	8698	2207	5850	972	9334	1879	<0.001
BMI (kg/m^2)	25.5	3.8	27.0	4.2	25.2	3.6	<0.001
18.5≤BMI<25 (%)	51		37.2		54.2		<0.001
25≤BMI<40 (%)	49		62.8		45.8		
Physical activity (MET-h/wk)	14.3	7.5, 24.2	12.0	4.9, 20.0	15.4	7.6, 24.2	<0.001 †
Alcohol intake (g/day)	4.9	1.1, 11.3	3.7	0.8, 8.5	5.2	1.3, 11.7	<0.001 †
Smoking (%)	19.3		21.1		18.9		0.09
Education (%)							0.02
Primary and secondary school	19.9		22.3		19.4		
Upper secondary school	39.8		41.3		39.5		
Academy/college/university (≤4y)	24.4		22.5		24.8		
Academy/college/university (>4y)	15.9		13.9		16.3		
Diseases (%)							
No disease	39.8		36.2		40.6		0.005
Total chronic disease ‡	55.5		59.2		54.7		0.005
Cardiovascular disease §	20.4		23.8		19.6		0.001
Diabetes ‖	3.0		4.8		2.6		<0.001
Chronic respiratory disease ¶	11.0		12.5		10.6		0.06
Cancer	6.2		6.9		6.0		0.29
Joint/muscle/skeletal disorder	36.4		39.1		35.8		0.04

P, percentiles; MET, metabolic equivalent task;
* Comparison of under- and plausible reporters: Two-sample t-test for continuous variables and chi-square test for categorical variables.
† Physical activity and alcohol were $\log_e$ transformed in the comparison of BMI groups.

[‡] Disease group composed of cardiovascular disease, diabetes, chronic respiratory disease, cancer and joint inflammation, muscle and skeletal disorder.
[§] Disease group composed of stroke, heart attack, angina and hypertension.
[||] Disease group composed of diabetes type 1 and type 2.
[¶] Disease group composed of asthma and chronic respiratory inflammation.

Table 2. Selected characteristics and prevalence of disease of all versus plausible reporters stratified by BMI
(Number and percentage; means and SD; medians and 25th and 75th percentiles (P))

	All reporters (*n* 6204)					Plausible reporters (*n* 5071)				
	18.5≤BMI<25		25≤BMI<40			18.5≤BMI<25		25≤BMI<40		
Characteristics	Mean or median	SD or P25, P75	Mean or median	SD or P25, P75	*P* *	Mean or median	SD or P25, P75	Mean or median	SD or P25, P75	*P* *
n	3167		3037			2746		2325		
Low energy reporters (%)	6.8		11.5		<0.001	-		-		
Age (years)	57.6	4.5	58.1	4.5	<0.001	57.6	4.5	58.2	4.5	<0.001
Energy intake (kJ/day)	8763	2235	8631	2176	0.02	9255	1960	9428	1774	0.001
BMI (kg/m^2)	22.7	1.6	28.5	3.0	-	22.6	1.6	28.2	2.8	-
Physical activity (MET-h/wk)	16.9	9.9, 26.6	12.5	7.2, 20.2	<0.001 †	17.5	10.5, 27.5	12.5	7.2, 20.5	<0.001 †
Alcohol intake (g/day)	5.7	1.6, 12.1	4.0	0.8, 10.4	<0.001 †	5.9	1.7, 12.4	4.3	0.8, 10.9	<0.001 †
Smoking (%)	21.3		17.2		<0.001	20.6		16.8		<0.001
Education (%)					<0.001					<0.001
Primary and secondary school	15.8		24.2			15.4		24.1		
Upper secondary school	38.6		41.2			38.5		40.8		
Academy/college/university (≤4y)	26.9		21.7			27.3		21.8		
Academy/college/university (>4y)	18.7		12.9			18.9		13.3		
Diseases (%)										
No disease	47.6		31.7		<0.001	47.4		32.6		<0.001
Total chronic disease ‡	20.8		38.1		<0.001	20.7		37.1		<0.001
Cardiovascular disease §	13.1		27.9		<0.001	13.1		27.3		<0.001
Diabetes ‖	1.7		4.8		<0.001	1.3		4.2		<0.001
Chronic respiratory disease ¶	8.6		13.1		<0.001	8.9		12.6		<0.001
Cancer	7.4		6.4		0.60	5.8		6.3		0.50
Joint/muscle/skeletal disorder	32.1		41.5		<0.001	31.5		40.9		<0.001

MET, metabolic equivalent task

* Comparison of BMI groups: Two-sample t-test for continuous variables and chi-square test for categorical variables.

† Physical activity and alcohol were $\log_e$ transformed in the comparison of BMI groups.

‡ Disease group composed of cardiovascular disease, diabetes, chronic respiratory disease, cancer and joint inflammation, muscle and skeletal disorder.

§ Disease group composed of stroke, heart attack, angina and hypertension.

‖ Disease group composed of diabetes type 1 and type 2.

¶ Disease group composed of asthma and chronic respiratory inflammation.

Table 3. Factor loadings* for the three dietary patterns found in the PCA for all (*n* 6204) and plausible (*n* 5071) reporters

	Prudent		Western		Continental	
Food group	All reporters	Plausible reporters	All reporters	Plausible reporters	All reporters	Plausible reporters
Vegetables	**0.65**	**0.64**	-0.05	-0.16	0.04	0.06
Fish, dinner	**0.53**	**0.57**	0.18	0.08	-0.09	-0.09
Fruits	**0.52**	**0.47**	0.03	-0.14	-0.20	-0.18
Herbs and spices	**0.50**	**0.38**	**-0.30**	**-0.41**	0.19	0.25
Berries	**0.48**	**0.46**	0.19	0.06	-0.16	-0.13
Nuts and seeds	**0.47**	**0.36**	-0.14	-0.28	0.05	0.08
Legumes	**0.40**	**0.34**	-0.22	-0.29	0.07	0.12
Meat dishes	**0.40**	**0.47**	0.22	0.18	0.16	0.16
Vegetarian food	**0.32**	0.20	-0.21	**-0.32**	0.004	0.07
Soup	**0.30**	**0.30**	0.13	0.05	0.09	0.12
Tea	**0.30**	0.22	-0.15	-0.23	0.01	0.05
Egg	0.24	0.24	-0.02	-0.05	0.12	0.12
Fish, breadspread	0.22	0.27	0.14	0.10	-0.04	-0.05
Water	0.21	0.24	-0.05	-0.06	-0.01	-0.01
Fruit juice	0.13	0.05	0.04	-0.04	0.02	0.03
Potatoes	-0.02	0.09	**0.59**	**0.57**	-0.11	-0.16
Sauce	-0.05	0.02	**0.57**	**0.58**	0.24	0.21
Refined grains	-0.04	-0.03	**0.54**	**0.50**	0.05	0.05
Processed meat	-0.02	0.03	**0.47**	**0.48**	**0.44**	**0.41**
Cakes and desserts	0.08	0.06	**0.46**	**0.40**	0.16	0.17
Margarine	-0.10	-0.10	**0.41**	**0.39**	0.06	0.04
Sweet spreads	0.06	0.07	**0.38**	**0.32**	-0.17	-0.16
Red meat and game	0.11	0.20	**0.37**	**0.40**	0.28	0.24
Wine	0.10	-0.01	**-0.32**	**-0.35**	**0.30**	0.29
Cheese	0.08	0.04	0.26	0.19	0.03	0.01
Whole grains	0.25	0.20	0.26	0.11	-0.19	-0.20
Mayonnaise	0.02	0.04	0.25	0.23	0.15	0.13
Sugar-sweetened beverages	-0.09	-0.09	0.23	0.21	-0.01	-0.01
Coffee	-0.11	-0.03	0.23	0.27	0.07	0.04
Butter	-0.01	-0.03	0.17	0.13	0.04	0.04
Sugar	-0.001	-0.03	0.14	0.09	0.02	0.02
Tomato sauce	0.21	0.15	0.09	0.05	**0.53**	**0.55**
Pasta	0.11	-0.02	-0.04	-0.11	**0.51**	**0.54**
Fat-rich potatoes	0.03	-0.003	0.21	0.19	**0.39**	**0.39**
Salty snacks	-0.08	-0.11	0.12	0.13	**0.38**	**0.37**
Pizza	-0.05	-0.07	0.14	0.14	**0.38**	**0.37**
Salad dressings	**0.36**	0.24	-0.15	-0.22	**0.37**	**0.42**
Rice	0.20	0.11	-0.05	-0.12	**0.36**	**0.40**
Poultry	**0.33**	**0.32**	-0.14	-0.16	**0.34**	**0.33**
Mustard	0.14	0.12	0.07	0.04	**0.34**	**0.34**
Sweets	0.02	-0.06	0.13	0.08	**0.30**	**0.30**
Soy sauce	0.25	0.17	-0.16	-0.22	0.28	**0.31**
Barbecue and taco seasoning	0.06	0.07	0.07	0.09	0.25	0.24
Low-fat dairy	0.06	-0.06	0.12	-0.02	-0.24	-0.24
Beer	0.05	-0.02	-0.03	-0.07	0.21	0.21
Sweeteners	-0.03	-0.01	0.05	0.09	0.19	0.17
High-fat dairy products	0.03	-0.09	0.17	0.06	-0.18	-0.17
Artificially sweetened beverages	-0.06	-0.03	0.02	0.06	0.16	0.14
Liquor	-0.05	-0.06	0.02	0.03	0.14	0.12
Total variance explained, %	6.1	5.2	5.9	6.0	5.4	5.5

PCA, principal component analysis
* Factor loadings with an absolute value ≥0.30 in bold font.

Table 4. Relation between prevalence of self-reported chronic disease and tertiles of dietary pattern scores among normal weight (18.5≤BMI<25) all and plausible reporters (Number; OR* and 95% CI)

	All reporters (*n* 3167)					Plausible reporters (*n* 2746)				
Disease	n	T1	T2	T3	*P*-trend	n	T1	T2	T3	*P*-trend
Total chronic disease [†]	1500					1303				
Western		1.00	1.11 (0.93, 1.33)	0.97 (0.78. 1.20)	0.87		1.00	1.12 (0.93 1.36)	1.04 (0.84, 1.28)	0.66
Continental		1.00	1.06 (0.89, 1.27)	0.87 (0.72, 1.05)	0.17		1.00	1.16 (0.96, 1.41)	0.88 (0.72, 1.07)	0.23
Prudent		1.00	0.95 (0.78, 1.15)	1.24 (1.00, 1.55)	0.05		1.00	1.07 (0.88, 1.30)	1.43 (1.14, 1.80)	0.002
Cardiovascular disease [‡]	416					359				
Western		1.00	1.31 (0.99, 1.73)	1.04 (0.74, 1.46)	0.72		1.00	1.36 (1.02, 1.82)	1.11 (0.79, 1.56)	0.41
Continental		1.00	1.22 (0.93, 1.60)	1.07 (0.79, 1.44)	0.60		1.00	1.33 (0.99, 1.78)	1.03 (0.74, 1.42)	0.75
Prudent		1.00	0.88 (0.66, 1.18)	1.15 (0.82, 1.62)	0.43		1.00	1.06 (0.79, 1.43)	1.60 (1.14, 2.26)	0.009
Diabetes [§]	43					36				
Western		1.00	1.33 (0.63, 2.79)	0.69 (0.25, 1.88)	0.58		1.00	1.08 (0.50, 2.37)	0.66 (0.25, 1.74)	0.45
Continental		1.00	2.14 (1.03, 4.41)	0.92 (0.36, 2.34)	0.89		1.00	2.42 (1.11, 5.27)	0.71 (0.23, 2.18)	0.92
Prudent		1.00	0.90 (0.40, 2.02)	1.54 (0.62, 3.81)	0.39		1.00	1.36 (0.59, 3.13)	2.07 (0.79, 5.44)	0.14
Chronic respiratory disease [‖]	281					245				
Western		1.00	1.09 (0.79, 1.49)	0.95 (0.64, 1.39)	0.83		1.00	1.32 (0.94, 1.84)	1.07 (0.73, 1.58)	0.60
Continental		1.00	1.04 (0.75, 1.42)	0.97 (0.70, 1.35)	0.85		1.00	1.30 (0.92, 1.82)	1.05 (0.73, 1.51)	0.79
Prudent		1.00	0.85 (0.60, 1.19)	1.13 (0.77, 1.67)	0.53		1.00	0.93 (0.66, 1.31)	1.26 (0.85, 1.86)	0.29
Cancer	191					160				
Western		1.00	1.24 (0.84, 1.81)	1.11 (0.70, 1.75)	0.61		1.00	1.08 (0.72, 1.62)	1.03 (0.65, 1.62)	0.88
Continental		1.00	1.16 (0.80, 1.66)	0.82 (0.55, 1.24)	0.39		1.00	1.27 (0.86, 1.88)	0.75 (0.47, 1.18)	0.27
Prudent		1.00	0.84 (0.57, 1.25)	0.91 (0.57, 1.44)	0.66		1.00	0.83 (0.55, 1.25)	1.09 (0.68, 1.75)	0.82
Joint/muscle/skeletal disorder	1000					865				
Western		1.00	1.08 (0.88, 1.32)	0.96 (0.75, 1.22)	0.78		1.00	1.04 (0.84, 1.29)	1.03 (0.81, 1.31)	0.79
Continental		1.00	1.03 (0.85, 1.26)	0.77 (0.62, 0.95)	0.02		1.00	1.11 (0.89, 1.37)	0.77 (0.62, 0.97)	0.04
Prudent		1.00	1.03 (0.83, 1.28)	1.44 (1.13, 1.85)	0.003		1.00	1.24 (1.00, 1.55)	1.69 (1.31, 2.18)	<0.001

OR, odds ratio; CI, confidence interval; T, tertile

* Adjusted for age (continuous), education (categorical), smoking (categorical), physical activity (continuous) and energy intake (continuous).

[†] Disease group composed of cardiovascular disease, diabetes, chronic respiratory disease, cancer and joint inflammation, muscle and skeletal disorder.

[‡] Disease group composed of stroke, heart attack, angina and hypertension.
[§] Disease group composed of diabetes type 1 and type 2.
[||] Disease group composed of asthma and chronic respiratory inflammation.

Table 5. Relation between prevalence of self-reported chronic disease and tertiles of dietary pattern scores among overweight/obese (25≤BMI<40) all and plausible reporters (Number; OR* and 95% CI)

	All reporters (*n* 3037)					Plausible reporters (*n* 2325)				
Disease	n	T1	T2	T3	*P*-trend	n	T1	T2	T3	*P*-trend
Total chronic disease [†]	1944					1470				
Western		1.00	0.86 (0.70, 1.05)	1.00 (0.79, 1.26)	0.97		1.00	0.96 (0.76, 1.21)	1.14 (0.89, 1.45)	0.28
Continental		1.00	0.88 (0.73, 1.08)	1.03 (0.84, 1.26)	0.77		1.00	0.87 (0.70, 1.10)	1.01 (0.79, 1.27)	0.95
Prudent		1.00	1.01 (0.83, 1.24)	1.45 (1.14, 1.84)	0.003		1.00	1.33 (1.06, 1.67)	1.46 (1.14, 1.87)	0.003
Cardiovascular disease [‡]	848					635				
Western		1.00	0.91 (0.71, 1.17)	1.05 (0.79, 1.40)	0.72		1.00	1.01 (0.75, 1.35)	1.29 (0.95, 1.76)	0.08
Continental		1.00	0.82 (0.65, 1.04)	1.03 (0.80, 1.33)	0.83		1.00	0.88 (0.67, 1.17)	1.12 (0.83, 1.49)	0.48
Prudent		1.00	1.12 (0.87, 1.43)	1.69 (1.26, 2.27)	0.001		1.00	1.49 (1.12, 1.97)	1.63 (1.19, 2.23)	0.002
Diabetes [§]	145					98				
Western		1.00	0.72 (0.46, 1.14)	0.90 (0.54, 1.52)	0.66		1.00	0.92 (0.52, 1.64)	1.27 (0.70, 2.28)	0.39
Continental		1.00	0.75 (0.48, 1.16)	0.95 (0.60, 1.51)	0.77		1.00	1.04 (0.60, 1.78)	1.30 (0.74, 2.28)	0.37
Prudent		1.00	1.59 (1.00, 2.54)	2.80 (1.62, 4.87)	<0.001		1.00	2.60 (1.40, 4.85)	3.82 (1.95, 7.51)	<0.001
Chronic respiratory disease [‖]	399					293				
Western		1.00	0.84 (0.62, 1.14)	1.00 (0.70, 1.41)	0.98		1.00	0.96 (0.67, 1.37)	1.15 (0.79, 1.67)	0.44
Continental		1.00	0.91 (0.68, 1.24)	1.18 (0.87, 1.61)	0.28		1.00	0.88 (0.62, 1.25)	1.27 (0.89, 1.82)	0.17
Prudent		1.00	0.94 (0.69, 1.26)	1.33 (0.93, 1.91)	0.15		1.00	1.55 (1.09, 2.22)	1.62 (1.09, 2.40)	0.02
Cancer	193					146				
Western		1.00	0.84 (0.55, 1.28)	1.34 (0.85, 2.10)	0.18		1.00	1.06 (0.64, 1.75)	1.68 (1.02, 2.77)	0.03
Continental		1.00	0.83 (0.55, 1.24)	1.23 (0.83, 1.83)	0.30		1.00	0.80 (0.51, 1.27)	1.11 (0.70, 1.76)	0.68
Prudent		1.00	0.69 (0.46, 1.03)	0.95 (0.60, 1.53)	0.71		1.00	0.91 (0.57, 1.45)	1.04 (0.63, 1.74)	0.89
Joint/muscle/skeletal disorder	1260					952				
Western		1.00	0.84 (0.67, 1.05)	1.00 (0.77, 1.29)	0.99		1.00	0.93 (0.72, 1.20)	1.09 (0.83, 1.43)	0.49
Continental		1.00	0.84 (0.68, 1.04)	0.96 (0.77, 1.20)	0.74		1.00	0.83 (0.65, 1.06)	0.92 (0.72, 1.19)	0.55
Prudent		1.00	1.01 (0.82, 1.26)	1.44 (1.11, 1.87)	0.007		1.00	1.24 (0.97, 1.60)	1.33 (1.01, 1.75)	0.04

OR, odds ratio; CI, confidence interval; T, tertile

* Adjusted for age (continuous), education (categorical), smoking (categorical), physical activity (continuous) and energy intake (continuous).

[†] Disease group composed of cardiovascular disease, diabetes, chronic respiratory disease, cancer and joint inflammation, muscle and skeletal disorder.

[‡] Disease group composed of stroke, heart attack, angina and hypertension.
[§] Disease group composed of diabetes type 1 and type 2.
[||] Disease group composed of asthma and chronic respiratory inflammation.

Supplemental Table 1. Food groupings used in dietary pattern analysis, 49 food groups.

Food group	Food items
Artificially sweetened beverages	Artificially sweetened soft drinks, artificially sweetened squash, artificially sweetened ice tea
Barbecue and taco seasoning	Taco seasoning, grill seasoning
Beer	Malt beer, pilsner, light beer
Berries	Blackberry, blueberry, raspberry, strawberry, cherry , cloudberry, rose hips, redcurrant, blackcurrant, cowberry
Butter	Butter, melted butter
Cakes and desserts	Dessert mousse, vanilla sauce, ice-cream, canned fruits, ice pop, muffins, chocolate cake, sponge cake with cream (with or without marzipan), sweet biscuits, Danish pastry, cinnamon bun/ sweet bun with vanilla cream
Cheese	Regular and low-fat Norwegian brown cheese, regular and low-fat hard cheese, regular and low-fat cheese spread, cream cheese
Coffee	Coffee, boiled/press, instant, espresso, café latte, cappuccino
Eggs	Eggs
Fat rich potatoes	Potato salad, potato gratin with cream, fried potatoes, homemade French fries, restaurant made French fries
Fish, dinner	Smoked salmon/brown trout, fried salmon/brown trout, sardine, herring, shrimps/crabs, fish cakes/fish pudding, fish balls, fish fingers, boiled cod/saithe/haddock/catfish/redfish, fried cod/saithe/haddock/ catfish/redfish, smoked mackerel, fried mackerel, fish gratin, wok with seafood and vegetables
Fish, bread spread	Caviar spread, roe paste, mackerel in tomato sauce
Fruit juice	Blueberry and aronia juice, cranberry and raspberry juice, rosehip and orange juice, orange juice, apple juice
Fruits	Apple, pear, banana, orange, clementine, peach/nectarine, kiwi, grapes, melon, pomegranate, fruit as spreads, fresh fruits salad, prune, raisins, other dried fruits
High-fat dairy products	High-fat milk, flavoured milk, sour cream, whipped cream, high-fat yoghurt
Herbs/spices	Dried and fresh basil, dried and fresh chili, dried and fresh oregano, dried and fresh thyme, cinnamon, cardamom, curry powder, black pepper powder, sweet red pepper powder, dried rosemary, garlic, fresh dill, fresh ginger, fresh peppermint, fresh parsley
Legumes	Legumes
Liquor	Spirits, cider, cocktail
Low-fat dairy products	Skimmed- and semi-skimmed milk, cultured/probiotic low-fat milk, low-fat yoghurt drink, low-fat yoghurt
Margarine	Normal and low-fat margarine
Mayonnaise	Regular and low fat remoulade/mayonnaise, regular and low fat spread with mayonnaise (Italian, shrimp etc.)
Meat dishes	Mutton and cabbage stew, stew with meat, vegetables and potatoes, wok with meat and vegetables
Mustard	Mustard
Nuts and seeds	Cashew nuts, peanuts, peanut butter, walnuts, hazelnuts, almonds, pecan nuts, pine nuts, pistachio nuts, sesame seeds, sunflower seeds
Pasta	Noodles, pasta, pasta with tomato sauce
Pizza	Pizza
Potatoes	Boiled, mashed
Poultry	Grilled chicken, chicken filet, chicken/turkey sausage, chicken/turkey grilled/wiener sausage
Processed meat	Meat balls, minced meat sauce, taco, kebab, lasagne, grilled/wiener sausage, minced meat sausage, bacon, regular and low fat liver paste, regular and low fat saveloy, salami
Read meat and game	Beef, roast of lamb/beef/pork, roast of game, hamburger, pork chops
Refined grains	White bread, crisp bread (wheat flour), hot dog bun, sweet muesli/breakfast cereal, pancake, rice porridge, waffle, sweet bun
Rice	Rice
Salad dressings	Thousand-island dressing, oil and vinegar dressing, pesto
Salty snacks	Potato chips, other salty snacks
Sauce	Brown/white sauce, béarnaise/hollandaise sauce
Soup	Tomato soup
Soy sauce	Soy sauce
Sugar	Pure sugar
Sugar-sweetened beverages	Fruit juices with added sugar, squash with sugar, sugar-sweetened soft drinks, iced tea with sugar
Sweets	Chocolate, dark chocolate, extra dark chocolate, sweets/jelly sweets, sweet pastille/candy
Sweeteners	Sweetener, sugar free pastille
Sweet spreads	Regular and low sugar jam, honey, chocolate/nut spread, other sweet spread
Tea	Black tea, green tea, herb tea
Tomato sauce	Ketchup, tomato sauce, taco sauce
Vegetables	Carrot, cabbage, swede, cauliflower, broccoli, brussels sprout, onion, spinach, sweet pepper, avocado, tomato, maize, frozen vegetables, mixed salad (with lettuce, cucumber, tomato and sweet pepper), vegetables as spread
Vegetarian food	Vegetarian food
Water	Tap water, mineral water
Wine	Red wine, white wine
Whole grains	Semi- and whole grain bread, crisp bread (whole meal), oat meal porridge and cereal, unsweetened muesli/breakfast cereal

Supplemental Table 2. Relation between prevalence of total chronic disease* and tertiles of dietary pattern scores divided in strata by BMI among the plausible reporters (OR and 95% CI)

		18.5≤BMI<25 (*n* 2746)			25≤BMI<40 (*n* 2325)		
	Model	T1	T2	T3	T1	T2	T3
Western	Crude	1.00	1.20 (1.00, 1.44)	1.10 (0.91, 1.33)	1.00	1.08 (0.86, 1.34)	1.38 (1.11, 1.72)
	1 (Age)	1.00	1.15 (0.96, 1.39)	1.09 (0.91, 1.32)	1.00	1.08 (0.86, 1.35)	1.42 (1.13, 1.77)
	2 (Model 1 + smoking)	1.00	1.15 (0.95, 1.38)	1.07 (0.88, 1.30)	1.00	1.08 (0.86, 1.35)	1.40 (1.12, 1.75)
	3 (Model 2 + education)	1.00	1.11 (0.92, 1.33)	0.99 (0.81, 1.21)	1.00	1.00 (0.79, 1.25)	1.25 (0.99, 1.57)
	4 (Model 3 + physical activity)	1.00	1.11 (0.92, 1.34)	1.00 (0.82, 1.22)	1.00	0.98 (0.78, 1.23)	1.21 (0.96, 1.52)
	5 (Model 4 + energy intake)	1.00	1.12 (0.93, 1.36)	1.04 (0.84, 1.28)	1.00	0.96 (0.76, 1.21)	1.14 (0.89, 1.45)
Continental	Crude	1.00	1.04 (0.86, 1.25)	0.70 (0.58, 0.85)	1.00	0.77 (0.62, 0.95)	0.77 (0.62, 0.95)
	1 (Age)	1.00	1.14 (0.94, 1.38)	0.84 (0.69, 1.03)	1.00	0.87 (0.70, 1.09)	1.01 (0.80, 1.27)
	2 (Model 1 + smoking)	1.00	1.15 (0.95 1.39)	0.84 (0.69, 1.02)	1.00	0.87 (0.70, 1.09)	1.01 (0.80, 1.27)
	3 (Model 2 + education)	1.00	1.16 (0.96, 1.40)	0.87 (0.71, 1.06)	1.00	0.89 (0.71, 1.11)	1.05 (0.83, 1.33)
	4 (Model 3 + physical activity)	1.00	1.16 (0.96, 1.41)	0.87 (0.71, 1.06)	1.00	0.88 (0.70, 1.11)	1.05 (0.83, 1.32)
	5 (Model 4 + energy intake)	1.00	1.16 (0.96, 1.41)	0.88 (0.72, 1.07)	1.00	0.87 (0.70, 1.10)	1.01 (0.79, 1.27)
Prudent	Crude	1.00	1.04 (0.86, 1.25)	1.28 (1.06, 1.54)	1.00	1.37 (1.10, 1.70)	1.48 (1.19, 1.83)
	1 (Age)	1.00	1.00 (0.83, 1.21)	1.23 (1.01, 1.49)	1.00	1.30 (1.04, 1.62)	1.39 (1.12, 1.73)
	2 (Model 1 + smoking)	1.00	1.01 (0.84, 1.22)	1.24 (1.03, 1.51)	1.00	1.30 (1.04, 1.62)	1.40 (1.13, 1.74)
	3 (Model 2 + education)	1.00	1.01 (0.83, 1.21)	1.25 (1.03, 1.52)	1.00	1.32 (1.05, 1.65)	1.44 (1.16, 1.79)
	4 (Model 3 + physical activity)	1.00	1.01 (0.83, 1.21)	1.25 (1.03, 1.52)	1.00	1.35 (1.08 1.69)	1.51 (1.21, 1.89)
	5 (Model 4 + energy intake)	1.00	1.07 (0.88, 1.30)	1.43 (1.14, 1.80)	1.00	1.33 (1.06, 1.67)	1.46 (1.14, 1.87)

OR, odds ratio; CI, confidence interval; T, tertile

* Disease group composed of cardiovascular disease, diabetes, chronic respiratory disease, cancer and joint inflammation, muscle and skeletal disorder