

Exploring dietary patterns, obesity and sources of bias: the Västerbotten Intervention Programme (VIP)

Maria Nyholm^{1,2,*}, Lauren Lissner¹, Agneta Hörnell³, Ingegerd Johansson⁴, Göran Hallmans⁵, Lars Weinehall⁵ and Anna Winkvist²

¹Department of Public Health and Community Medicine/Public Health Epidemiology, University of Gothenburg, BOX 454, SE-405 30 Göteborg, Sweden: ²Department of Clinical Nutrition, University of Gothenburg, Göteborg, Sweden: ³Department of Food and Nutrition, Umeå University, Umeå, Sweden: ⁴Department of Odontology, Umeå University, Umeå, Sweden: ⁵Department of Public Health and Clinical Medicine, Umeå University, Umeå, Sweden

Submitted 17 October 2011: Final revision received 30 March 2012: Accepted 21 May 2012: First published online 5 July 2012

Abstract

Objective: Dietary patterns capture the overall diet and thereby provide information on how nutrients are consumed in combinations, and have been suggested to be a better method than studying single nutrients. The present study explored the relationship between dietary patterns at baseline and incidence of obesity at 10-year follow-up in women.

Design: A longitudinal study using baseline measurements from 1992–1996, including food intake, medication, heredity, socio-economic status, lifestyle and measured body composition, and follow-up data collected in 2002–2006 including measured body composition.

Setting: Data from the Västerbotten Intervention Programme (VIP) in Sweden.

Subjects: A total of 6545 initially non-obese women aged 30–50 years.

Results: Among women reporting plausible energy intakes, the 'Fruit and vegetables cluster' predicted the highest incidence of obesity (OR = 1.76, 95% CI 1.11, 2.76; $P = 0.015$) compared with women in the other food pattern groups combined. When adjusting for metabolic factors and BMI at baseline, the risk for obesity in the 'Fruit and vegetables cluster' was attenuated to non-significance. In contrast, high intake of fruit *per se* was associated with a decreased risk of developing obesity (OR = 0.69, 95% CI 0.51, 0.91; $P = 0.010$).

Conclusions: Dietary pattern groups identified by cluster analysis are likely to reflect characteristics in addition to diet, including lifestyle, previous and current health status and risk factors for future disease, whereas intake of fruit *per se* was a stable indicator and less affected by baseline characteristics. These results underscore the need for complementary methods in understanding diet–disease relationships.

Keywords
Dietary patterns
Fruit and vegetables
Incidence of obesity
Bias
Biomedical risk factors
Socio-economic status
Lifestyle

Obesity is a major public health issue as it contributes to CVD, type 2 diabetes, cancer and low quality of life^(1–3). Lifestyle factors such as physical inactivity and diet contribute to the development of obesity^(2,4). The relationship between diet and obesity is complex and difficult to estimate, due to the low accuracy of dietary data collected among free-living persons^(5,6). The methodological approach to describing the association between diet and diseases has often focused on a single nutrient or specific nutrients^(7,8). Analysis of dietary patterns has been suggested to be a better approach for studying diet–disease relationships, as dietary patterns define the overall diet and thereby provide information on how nutrients are consumed in combinations^(9–11).

Prospective studies have found associations between a low intake of fruit and vegetables and an increase in weight, BMI and waist-to-hip ratio^(8,12). One advantage of

prospective, as opposed to cross-sectional, studies is that reported dietary intakes are less likely to be biased by the knowledge of outcome⁽¹³⁾. However, biases may still exist that could confound the diet–disease association even with prospective designs. For example, misreporting of food intake and past dietary modifications might be important confounders in diet and disease relationships also in prospective studies^(14,15).

The largest ongoing population-based prospective study of diet and anthropometry measurements in Sweden is the Västerbotten Intervention Programme (VIP). Between 1992 and 2005, a total of 60 000 men and women were recruited. Four dietary patterns have recently been identified among the women and three among the men. In the present study of initially non-obese women, we explore the relationship between the VIP dietary patterns at baseline and incidence

of obesity at 10-year follow-up. Furthermore, we hypothesize that biases in dietary reporting at baseline may distort the relationship between diet pattern and body weight change.

Materials and methods

Study base

Since 1985, all inhabitants in Västerbotten county in northern Sweden have been invited by their primary health-care centre for a health check-up and to answer a diet and lifestyle questionnaire, when they turn 30, 40, 50 and 60 years of age. The annual participation rates on average have been about 60% of the invited population, and only marginal differences in background characteristics between participants and non-participants have been found⁽¹⁶⁾. The assessment of dietary intake was initially performed with an FFQ containing eighty-four food items; however, from 1996 a version with sixty-four food items also was introduced⁽¹⁷⁾. Among participants aged 40–60 years 40.8% had filled in the eighty-four-item version, and among the 30-year-old participants 83.7% had filled in the eighty-four-item version. Furthermore, data collected before 1992 were not optically readable and therefore were not included in the present study⁽¹⁷⁾.

The present study includes data collected between 1992 and 2006 on women between 30 and 50 years of age who had their first health check-up during the years 1992–1996, who had filled in the eighty-four-item FFQ and who had a follow-up in the years 2002–2006. The VIP subsample used in the present study thus consisted of 8560 women. The project was approved by the regional Ethics Committee, University of Gothenburg, Sweden.

Diet intake assessment and food intake level

Dietary intake was assessed with the eighty-four-item self-administered FFQ and frequency of food intake was reported on a nine-level scale from none to ≥ 4 times daily^(17–19). Portion sizes were indicated by colour photographs of four plates with a gradient of increasing portion size for staple food, meat and vegetables⁽²⁰⁾. Daily intakes were calculated by multiplying frequency of intake by a portion size value using the national food composition database^(19–21). Energy and nutrient contents were calculated using the software MATs (Rudans Lättdata 1994, Västerås Sweden). Food intake level (FIL), i.e. the ratio between reported total energy intake and estimated BMR, was used to assess the accuracy of diet intake data⁽²²⁾.

Dietary patterns have previously been assessed within the whole VIP cohort using an a posteriori approach by *k*-means cluster analysis⁽¹⁷⁾. Four mutually exclusive dietary clusters were identified among the women and labelled: 'High fat' (high intake of high-fat spreads and high-fat milk products); 'Tea and ice cream' (high consumption of tea and ice cream); 'Coffee and sandwich'

(high consumption of coffee, low-fat spread and cold cuts); and 'Fruit and vegetables' (high consumption of fruit, vegetables, chicken, fish, red meat and boiled potatoes)^(17,18). Each woman was classified to one of the four identified patterns.

A *post hoc* hypothesis was that fruits and vegetables *per se* were related to obesity development. To test this, total fruit and vegetable intakes were extracted from the FFQ and the response to each food item was converted into an average daily intake for each individual participant. All intakes of fruit (including berries) and vegetable items were summarized into one fruit intake variable and one vegetable intake variable. High intake *per se* was defined as above the 75th percentile, i.e. a fruit intake of 1.34 items/d and a vegetable intake of 0.84 items/d⁽¹⁷⁾.

Anthropometry

Body weight and standing height were measured at the health centre. BMI was categorized into subgroups according to WHO definitions (underweight: BMI < 18.50 kg/m², normal weight: BMI = 18.51–25.00 kg/m², overweight: BMI = 25.01–29.99 kg/m², obesity: BMI \geq 30.00 kg/m²)⁽²³⁾. In evaluating the incidence of obesity, women were dichotomized into those who developed obesity or not.

Analytical sample

Women were excluded from further analysis if at baseline they were defined as underweight or obese (*n* 1117), with missing FIL or unrealistic FIL levels (values below the 5th percentile or above the 97.5th percentile of the FIL distribution; *n* 539), with missing data on dietary pattern (*n* 230) or with missing data on level of physical activity (*n* 129).

A total of 6545 women with acceptable data on portion sizes, intake frequencies and FIL were included. The mean duration of follow-up was 9.9 years and age at baseline was between 28.7 and 51.8 years.

Identification of low and high energy reporters

FIL was compared with physical activity level (PAL), which equals the ratio of total energy expenditure to BMR⁽²⁴⁾. PAL was estimated from self-reported level of physical activity at work and at leisure time⁽²⁵⁾. Physical activity at work was defined according to how the women described activity level at work and categorized as follows: level 1 = sedentary (sitting still); level 2 = very light (work that comprises very light activities); level 3 = light (work that comprises light activities) and moderate (postman on bicycle); and level 4 = heavy (heavy industrial work). Level 1 and level 2 were jointly considered as very light work activity. Physical activity at leisure time was defined according to how often physical activity was completed and categorized into five categories: never, not regularly, regularly 1–2 times/week, regularly 2–3 times/week and regularly >3 times/week.

Women with FIL below the lower 95% confidence interval limit of the calculated PAL were defined as low energy reporters, women within the confidence interval limits were defined as adequate energy reporters and women above the upper 95% confidence interval limit were defined as high energy reporters. The latter group was small and therefore combined with the adequate energy reporters.

Social variables

Age was treated as a continuous variable. Length of education in years was categorized into low education (<13 years) and high education (≥ 13 years). Cohabiting status was categorized as living single (cohabiting with children was categorized as single) or as married/cohabiting. Living area was categorized in two categories: living in area with $\leq 15\,000$ or $>15\,000$ inhabitants. Alcohol intake was divided into: abstainers (<5 g alcohol daily) and users (≥ 5 g alcohol daily). Current smokers were categorized as smokers, and never smokers/previous smokers were considered as non-smokers. Women who participated in leisure-time physical activity more than once weekly were defined as physically active and those reporting less were considered to have a sedentary lifestyle.

Biomedical risk factors

Blood specimens were obtained from the participants for standard measurements of serum lipids and capillary blood was used for a glucose tolerance test. Serum lipids, blood glucose and blood pressure were categorized as high or low levels according to the guidelines of the Swedish Medical Products Agency for prevention of CVD⁽²⁶⁾. Increased serum cholesterol was defined as ≥ 6.5 mmol/l, increased serum TAG was defined as ≥ 1.7 mmol/l. Hyperglycaemia was defined as fasting blood glucose ≥ 7.0 mmol/l or 2 h post-test blood glucose value ≥ 11.0 mmol/l^(27,28). High blood pressure was defined as systolic blood pressure ≥ 140 mmHg and/or diastolic blood pressure ≥ 90 mmHg.

Information about heredity was collected by the following questions at the health check: 'Have either of your parents or siblings developed cardiovascular heart disease or stroke before the age of 60?' and 'Are either of your parents or siblings suffering from type 2 diabetes?'.

Furthermore, the women also reported their intake of medications for high blood pressure, CVD and high lipids.

Statistical analysis

Differences in baseline characteristics between those who did and did not develop obesity were evaluated using Student's *t* test and the χ^2 test. Differences among the four dietary patterns in the incidence of obesity were investigated using the χ^2 test. For further analyses the four dietary patterns were re-grouped into two patterns: 'Fruit and vegetables cluster' and 'Mixed pattern group'

('High fat', 'Tea and ice cream' and 'Coffee and sandwich' combined), to explore a 'healthy dietary pattern' with the other patterns.

Multivariate logistic regression models were used to estimate the associations between dietary intake ('Fruit and vegetables cluster', 'Mixed pattern group', high intake of fruit or vegetables) and the risk of developing obesity as odds ratios with 95% confidence intervals. In the regression models estimation high intake of fruit or vegetables *per se* was defined according to the fourth quartile of intake. Three multivariate models were then used. The first model included age, socio-economic factors (education, cohabitation status and living area) and lifestyle factors (smoking, alcohol and leisure-time activity) as covariates. The second model included age, baseline BMI, heredity for CVD, heredity for type 2 diabetes, medication (for CVD, high lipids, high blood pressure) and biomedical risk factors (high total cholesterol, high TAG, high blood pressure, hyperglycaemia) as covariates. The final model included all covariates mentioned above.

All analyses were stratified for adequacy of energy reporting (low or adequate). All tests were two-sided and significance was assumed at $P < 0.05$. The statistical analyses were performed using the SPSS statistical software package version 16.0 (SPSS Inc.).

Results

At baseline the mean BMI was 23.7 (SD 2.6) kg/m²; after 10 years of follow-up 535 (8.2%) of the initially non-obese women had developed obesity, of whom 490 (7.5%) were overweight at baseline. The mean baseline BMI was significantly higher ($P < 0.001$) and FIL was significantly lower ($P = 0.009$) in women who later developed obesity, compared with women who did not develop obesity (Table 1). At baseline, the women who later developed obesity were more often current smokers ($P = 0.017$) and had more heredity for CVD ($P = 0.005$) and type 2 diabetes ($P < 0.001$). Further, women who later developed obesity had a higher proportion of biomedical risk factors ($P = 0.002$, $P < 0.001$ and $P < 0.001$ for high cholesterol, TAG and blood pressure, respectively) and were more often taking medications for hypertension already at baseline ($P < 0.001$).

When analysing all women together, no significant differences in the incidence of obesity were found among the four dietary patterns at baseline ($P = 0.282$). The three groups 'High fat', 'Coffee and sandwich' and 'Tea and ice cream' showed similar obesity outcome; these groups were therefore combined into a single 'Mixed pattern group', which differed marginally from the 'Fruit and vegetables cluster' ($P = 0.076$; Fig. 1). However, when stratifying for misreporting among adequate energy reporters, women in the 'Fruit and vegetables cluster' at baseline had the highest

Table 1 Baseline characteristics in relation to incidence of obesity: initially non-obese women aged 30–50 years participating in the Västerbotten Intervention Programme (VIP)

Baseline characteristic	BMI group after 10 years of follow-up						P value
	All (<i>n</i> 6545)		BMI < 30 kg/m ² (<i>n</i> 6010)		BMI ≥ 30 kg/m ² (<i>n</i> 535)		
	Mean	SD	Mean	SD	Mean	SD	
BMI (kg/m ²)	23.7	2.6	23.4	2.4	27.4	1.7	<0.001
Food intake level (FIL)	1.32	0.30	1.33	0.30	1.24	0.28	0.009
	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	
Lifestyle factors							
Current smoker	11768	27.2	1600	26.8	168	31.6	0.017
Alcohol intake ≥ 5.0 g/d	1075	16.4	1016	16.9	59	11.0	<0.001
Physically active	3999	61.1	3702	61.2	297	55.5	0.006
Socio-economic factors							
Higher education	1932	29.6	1809	30.2	123	23.0	0.001
Single	815	12.5	732	12.2	83	15.5	0.025
Living area with ≤15 000 inhabitants	4373	66.8	4045	67.3	328	61.3	0.005
Heredity							
Parents with CVD	1196	18.4	1074	18.0	123	23.3	0.005
Parents with type 2 diabetes	1092	16.8	969	16.2	122	23.0	<0.001
Medication for							
Hypertension	231	3.6	191	3.3	40	7.8	<0.001
CVD	23	0.4	21	0.4	2	0.4	0.913
High cholesterol	4	0.1	4	0.1	0	0.0	0.553
Metabolic risk factors							
High cholesterol	964	14.8	861	14.4	103	19.4	0.002
High TAG	708	11.7	617	11.1	91	17.9	<0.001
High blood pressure	863	13.4	751	12.7	112	21.5	<0.001
Hyperglycaemia	68	1.1	59	1.0	9	1.7	0.132
Fruit and vegetable intake (items/d)							
High intake of vegetables	1259	19.2	1153	19.2	106	19.9	0.710
High intake of fruits	1766	27.0	1649	27.4	117	21.9	0.006
Low energy reporters	2693	41.1	2423	40.3	270	50.4	<0.001

incidence of obesity compared with women in the 'Mixed pattern group' ($P = 0.015$; Fig. 1).

Among adequate and low reporters combined, the overall risk of developing obesity among women in the 'Fruit and vegetables cluster' compared with the 'Mixed pattern group' was $OR = 1.32$ (95% CI 0.97, 1.79; $P = 0.076$; Table 2). This marginally elevated risk was not due to misreporting of energy, as it was statistically significant only among adequate energy reporters ($OR = 1.76$, 95% CI 1.11, 2.76; $P = 0.015$; Table 2).

In a multivariate model adjusting for differences in age, baseline BMI, heredity, medication and biomedical risk factors, the risk of developing obesity in the 'Fruit and vegetables cluster' compared with the 'Mixed cluster group' in adequate energy reporters was attenuated to $OR = 1.43$ (95% CI 0.78, 2.62; $P = 0.242$; Table 2). This was true also in the final model in adequate energy reporters adjusting for age, baseline BMI, heredity, medication and biomedical risk factors, socio-economic factors and lifestyle factors ($OR = 1.47$, 95% CI 0.80, 2.69; $P = 0.212$; Table 2).

Among women in the 'Fruit and vegetables cluster' (adequate energy reporters), 62% of the women exhibited a high fruit intake *per se* and 81% exhibited a high vegetable intake *per se*. In Table 3, the relationships

between high daily intakes of fruit and vegetables *per se* and incidence of obesity are shown. There was no significant association between daily intake of vegetables *per se* and the risk of developing obesity, either in low energy reporters ($OR = 1.04$, 95% CI 0.73, 1.46; $P = 0.850$) or in adequate energy reporters ($OR = 1.12$, 95% CI 0.84, 1.51; $P = 0.435$). These results were not affected by further multivariate adjustment (Table 3). Similarly, a high daily intake of fruit *per se* was not significantly associated with obesity among low energy reporters ($OR = 0.97$, 95% CI 0.70, 1.33; $P = 0.842$). However, among adequate energy reporters daily fruit intake showed a significant protective effect on obesity ($OR = 0.69$, 95% CI 0.51, 0.91; $P = 0.010$), which remained significant in a final multivariate model adjusting for age, baseline BMI, heredity, medication and biomedical risk factors, socio-economic factors and lifestyle factors ($OR = 0.67$, 95% CI 0.47, 0.97; $P = 0.035$; Table 3).

Discussion

An unexpected association between the 'Fruit and vegetables cluster' and higher incidence of obesity was initially seen in the group judged to be adequate energy reporters

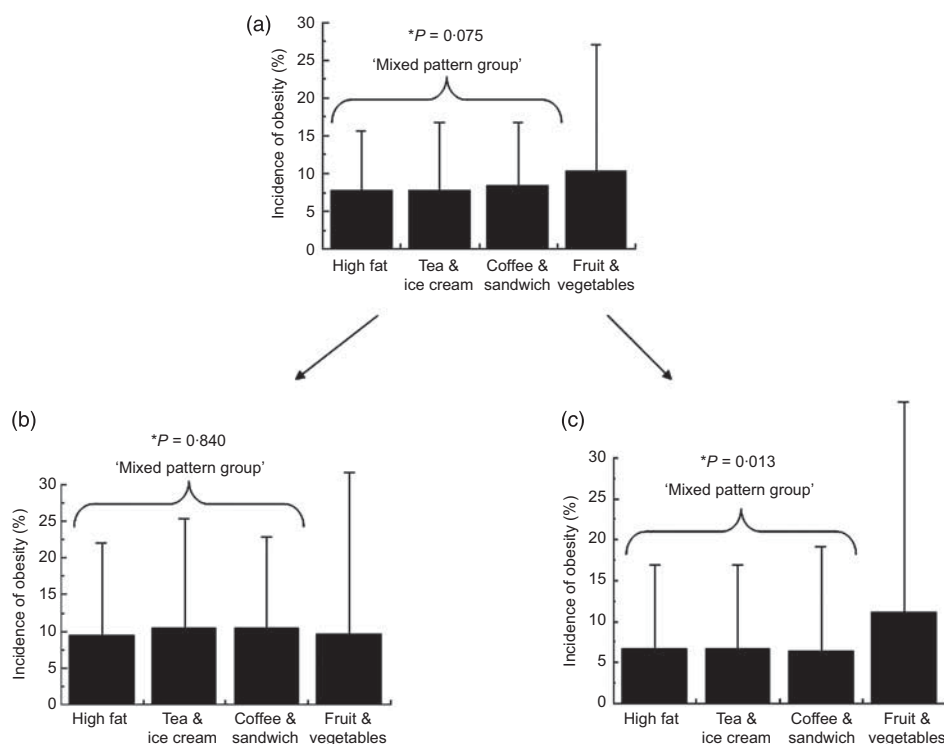


Fig. 1 Incidence of obesity at 10-year follow-up according to dietary pattern among initially non-obese women (n 6545) aged 30–50 years participating in the Västerbotten Intervention Programme (VIP): (a) all women; (b) low energy reporters; (c) adequate energy reporters. Values are means with 95 % confidence intervals represented by vertical bars. * P for the difference in obesity incidence between the 'Fruit and vegetables cluster' v. the 'Mixed pattern group' (χ^2 test)

Table 2 The odds for developing obesity at 10-year follow-up among women in the 'Fruit and vegetables cluster' compared with women in the 'Mixed pattern group' ('High fat', 'Tea and ice cream' and 'Coffee and sandwich' combined): initially non-obese women (n 6545) aged 30–50 years participating in the Västerbotten Intervention Programme (VIP)

	OR*	95 % CI	P value
All (n 6545)			
Crude	1.32	0.97, 1.79	0.076
Socio-economic factors and lifestyle factor†	1.35	0.99, 1.84	0.056
CVD risk profile‡	1.20	0.82, 1.78	0.351
All factors§	1.30	0.89, 1.91	0.174
Low energy reporters (n 2694)			
Crude	0.96	0.64, 1.45	0.840
Socio-economic factors and lifestyle factor†	1.02	0.67, 1.55	0.930
CVD risk profile‡	1.01	0.60, 1.69	0.966
All factors§	1.09	0.65, 1.82	0.756
Adequate energy reporters (n 3851)			
Crude	1.76	1.11, 2.76	0.015
Socio-economic factors and lifestyle factor†	1.78	1.13, 2.82	0.013
CVD risk profile‡	1.43	0.78, 2.62	0.242
All factors§	1.47	0.80, 2.69	0.212

*The odds ratio, with 95 % confidence interval, for developing obesity, estimated using logistic regression analysis. 'Mixed pattern group' was used as the reference.

†Adjusted for age, lifestyle and socio-economic factors.

‡Adjusted for age, heredity, medication, biomedical risk factors and baseline BMI.

§Adjusted for all factors in the above models.

at baseline. Specifically, the 'Fruit and vegetables cluster' had a higher risk to develop obesity after 10 years of follow-up compared with women in the 'Mixed pattern group' ('High fat', 'Coffee and sandwich', 'Tea and ice cream' combined). However, the relationship became

attenuated when adjusting for biomedical risk factors at baseline in multivariate models. In exploring fruit and vegetable intake further, high daily intakes of fruit and vegetables *per se* were not related to the risk of developing obesity. Rather, among adequate dietary reporters,

Table 3 The odds for developing obesity at 10-year follow-up among women with high intakes of fruits and vegetables (according to the fourth quartile of intake *per se*) compared with women with low intakes: initially non-obese women (*n* 6545) aged 30–50 years participating in the Västerbotten Intervention Programme (VIP)

	OR*	95% CI	P
Low energy reporters			
High intake of vegetables			
Crude	1.04	0.73, 1.46	0.850
Socio-economic factors and lifestyle factor†	1.11	0.79, 1.58	0.545
CVD risk profile‡	0.94	0.60, 1.45	0.763
All factors§	0.94	0.60, 1.46	0.766
High intake of fruits			
Crude	0.97	0.70, 1.33	0.842
Socio-economic factors and lifestyle factor†	1.01	0.73, 1.40	0.940
CVD risk profile‡	1.07	0.71, 1.61	0.745
All factors§	1.07	0.71, 1.62	0.734
Adequate energy reporters			
High intake of vegetables			
Crude	1.12	0.84, 1.51	0.435
Socio-economic factors and lifestyle factor†	1.18	0.88, 1.59	0.278
CVD risk profile‡	0.93	0.64, 1.34	0.683
All factors§	0.94	0.65, 1.37	0.751
High intake of fruits			
Crude	0.69	0.51, 0.91	0.010
Socio-economic factors and lifestyle factor†	0.68	0.51, 0.91	0.010
CVD risk profile‡	0.66	0.46, 0.96	0.028
All factors§	0.67	0.47, 0.97	0.035

*The odds ratio, with 95% confidence interval, for developing obesity, estimated using logistic regression analysis.

†Adjusted for age, socio-economic factors and lifestyle factors.

‡Adjusted for age, heredity, medication, biomedical risk factors and baseline BMI.

§Adjusted for all factors included in models above.

a high daily intake of fruit had a protective association with obesity and the association was robust even after adjusting for age, baseline BMI, heredity, medication and biomedical risk factors at baseline. These results add support to the need for complementary methods, beyond simply adjusting for age and lifestyle and socio-economic factors, in interpreting the diet–disease relationship.

The present study may be contrasted with other prospective studies that have reported an inverse association, i.e. between dietary patterns characterized by high fibre intake and high intake of fruit and vegetables and annual negative change in BMI^(29,30). However, in our study women who later developed obesity had a higher proportion of biomedical risk factors and of taking medications for hypertension already at baseline. Women at risk of gaining further weight might already have modified their diets due to relatively high weight, medication and biomedical risk factors at baseline. Another explanation might be that patients with risk factors for the metabolic syndrome, such as hypertension, are more exposed to information and education on healthy food intakes at their health-care centre. These women may have given a socially desirable response when reporting diet intake in order to respond within expected standards^(31,32).

The risk of developing obesity with the 'Fruit and vegetables cluster' in the present study was not stable after adjusting for biomedical risk factors and some other factors at baseline. Likewise, Togo *et al.* showed that the inverse association between a pattern of high fruit and vegetable intake and obesity disappeared after adjusting

for physical activity, smoking cessation and childbirth at baseline⁽³³⁾. These results might indicate that cluster analysis based on dietary groups captures characteristics additional to those of dietary intake, such as family-related factors⁽³¹⁾. High daily fruit intake *per se* was, however, a stable indicator for not developing obesity after 10 years in our study. The inverse associations between high intake of fruit and obesity have been shown in other prospective studies not based on pattern analysis^(12,34). Finally, cluster analysis is not well suited to analyse changes in diet over time. This is because clusters are generated inductively, meaning that different sets of clusters may be generated at different time points. This makes it challenging to classify diet changes over time. Methodological approaches including both dietary patterns and intakes of whole food items might be needed to avoid contradictory results in epidemiological studies.

Strengths of the present study were the prospective design with a large sample size and a follow-up time as long as 10 years. The four dietary patterns used in the study have been defined among over 30 000 women from Västerbotten^(17,18). However, the use of self-reported diet intakes is subject to measurement errors and bias, and low energy reporters are a common problem in dietary studies^(15,25). In our study the proportion of low energy reporters was high (58.9%) compared with other studies^(15,33). However, we used a strict definition for under-reporting by comparing individuals' estimated PAL (including both physical activities at work and leisure time) with FIL and this might have contributed to the higher

proportion of low energy reporters in our study. Furthermore, an FFQ with only eighty-four food items was used in our study and this might also have contributed to the high proportion of low energy reporters.

Other studies have shown that another source of bias in prospective studies concerning women and future body weight gain is gestational weight gain; however, such data were not available within the VIP study⁽³⁵⁾.

Conclusions

An unexpected association between fruit and vegetable intake and incidence of obesity was found in the present study, in which the clusters predicted more obesity whereas the food groups did not. Among accurate energy reporters, non-obese women characterized as having a dietary pattern with high intakes of fruits and vegetables were most likely to become obese after 10 years. Concurrently, high intake of fruit *per se* had a protective effect on the risk of developing obesity, whereas no such trend was seen for high intake of vegetables. However, low energy reporting was not the explanation for the unexpected result. Rather, the 'Fruit and vegetables cluster' based on cluster analysis was associated with the women's baseline characteristics, whereas intake of fruit *per se* was a stable indicator and less affected by baseline characteristics. We conclude that food pattern analysis should be used within caution in studying the diet-obesity association, because such definitions of dietary behaviour may introduce characteristics that are not solely related to the food of interest.

Acknowledgements

The study was supported by grants from the Swedish Council for Working Life and Social Research: the Swedish Cancer Society, the Swedish Research Council, the Wallenberg Foundation and Västerbotten County Council, Sweden. The authors declare that there are no competing financial interests in relation to this work and no conflict of interest. M.N., L.L. and A.W. were responsible for conception, design, statistical analyses and interpretation of the data. M.N. wrote the first manuscript draft. All authors critically reviewed and contributed to the final draft (M.N., L.L., A.W., I.J., A.H., G.H. and L.W.). All authors had full access to the data (M.N., L.L., A.W., I.J., A.H., G.H. and L.W.). The authors gratefully acknowledge the participants from the Västerbotten Intervention Programme.

References

1. Visscher TLS & Seidell JC (2001) The public health impact of obesity. *Annu Rev Public Health* **22**, 355–375.
2. Nyholm M, Gullberg B, Haglund B *et al.* (2008) Higher education and more physical activity limit the development of obesity in a Swedish rural population. The Skaraborg project. *Int J Obes (Lond)* **32**, 533–540.
3. Hill JO (2006) Understanding and addressing the epidemic of obesity: an energy balance perspective. *Endocr Rev* **27**, 750–761.
4. Hu FB, Li TY, Colditz GA *et al.* (2003) Television watching and other sedentary behaviours in relation to risk of obesity and type 2 diabetes mellitus in women. *JAMA* **289**, 1785–1791.
5. Tarasuk VS & Brooker A-S (1997) Interpreting epidemiologic studies of diet-diseases relationship. *J Nutr* **127**, 1847–1852.
6. Moeller SM, Reedy J, Millen AE *et al.* (2007) Dietary patterns: challenges and opportunities in dietary pattern research. An experimental biology workshop, 1 April 2006. *J Am Diet Assoc* **107**, 1233–1239.
7. Koh-Banerjee P, Chu NF, Spiegelmann D *et al.* (2003) Prospective study of the association of changes in dietary intake, physical activity, alcohol consumption, and smoking with 9-y gain in waist circumference among 16 587 US men. *Am J Clin Nutr* **78**, 719–727.
8. Halkjær J, Tjønnelund A, Overvad K *et al.* (2009) Dietary predictors of 5-year changes in waist circumference. *J Am Diet Assoc* **109**, 1356–1366.
9. Quatromoni PA, Copenhaver DL, D'Agostino RB *et al.* (2002) Dietary patterns predict the development of overweight in women: the Framingham Nutrition Studies. *J Am Diet Assoc* **102**, 1240–1246.
10. Berg CM, Lappas G, Strandhagen E *et al.* (2008) Food patterns and cardiovascular disease risk factors: the Swedish INTERGENE research program. *Am J Clin Nutr* **88**, 289–297.
11. Román-Viñas B, Barba LR, Ngo J *et al.* (2009) Validity of dietary patterns to assess nutrient adequacy. *Br J Nutr* **101**, 12–20.
12. Vioque J, Weinbrenner T, Castelló A *et al.* (2008) Intake of fruits and vegetables in relation to 10-years weight gain among Spanish adults. *Obesity (Silver Spring)* **16**, 664–670.
13. Willett W (1990) *Nutritional Epidemiology*. Oxford: Oxford University Press.
14. Wirfält 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 Malmö Diet and Cancer study. *Public Health Nutr* **3**, 159–173.
15. Mattisson I, Wirfält E, Aronsson AC *et al.* (2005) Misreporting of energy: prevalence, characteristics of misreport and influence on observed risk estimates in Malmö Diet and Cancer cohort. *Br J Nutr* **94**, 832–842.
16. Weinehall L, Hallgren C-G, Westman G *et al.* (1998) Reduction of selection bias in primary prevention of cardiovascular disease through involvement of primary health care. *Scand J Prim Health Care* **16**, 171–176.
17. Winkvist A, Hörnell 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.
18. Winkvist A, Hörnell 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.
19. Johansson I, Hallman G, Wikman Å *et al.* (2002) Validation and calibration of food frequency questionnaire measurements in the Northern Sweden Health and Disease cohort. *Public Health Nutr* **5**, 487–496.
20. Bergström L, Kylberg E, Hagman U *et al.* (1991) The food composition data base system (KOST-systemet) – its use for nutrient values. *Vår föda* **43**, 439–447.
21. Livsmedelsverket (1999) *Vikttabeller (Weight Tables)*. Uppsala: Livsmedelsverket.
22. Schofield WN, Schofield C & James WPT (1985) Basal metabolic rate. *Hum Nutr Clin Nutr* **39**, Suppl. 1, 1–96.
23. World Health Organization (1998) *Obesity: Prevention and Managing the Global Epidemic. Report of a WHO*

- Consultation on Obesity, Geneva, 3–5 June 1997. WHO/NUT/NCD/98.1.* Geneva: WHO.
24. Goldberg GR, Black AE, Jebb SA *et al.* (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**, 568–581.
 25. Johansson G & Westerterp KR (2008) Assessment of the physical activity level with two questions: validation with doubled labelled water. *Int J Obes (Lond)* **32**, 1031–1033.
 26. Läkemedelsverket (2006) Förebyggande av aterosklerotisk hjärt-kärlsjukdom (Prevention of cardiovascular disease). *Information från Läkemedelsverket (Information from the Swedish Medical Products Agency)* 3:2006, 16–31.
 27. Norberg M, Wall S, Boman K *et al.* (2010) The Västerbotten Intervention Programme: background, design and implications. *Glob Health Action* **3**, 10.
 28. World Health Organization (2006) *Definition and Diagnosis of Diabetes Mellitus and Intermediate Hyperglycemia*. Geneva: WHO; available at http://whqlibdoc.who.int/publications/2006/9241594934_eng.pdf
 29. Newby PK, Muller D, Hallfrisch J *et al.* (2004) Food patterns measured by factor analysis and anthropometric changes in adults. *Am J Clin Nutr* **80**, 504–513.
 30. Sánchez-Villegas A, Bes-Rastrollo M, Martínez-González MA *et al.* (2006) Adherence to a Mediterranean dietary pattern and weight gain in a follow-up study: the SUN cohort. *Int J Obes (Lond)* **30**, 350–358.
 31. Hérbert JR, Peterson KE, Hurley TG *et al.* (2001) The effect of social desirability trait on self-reported dietary measures among multi-ethnic female health centre employees. *Ann Epidemiol* **11**, 417–427.
 32. Rennie KL, Sirevo M & Jebb SA (2006) Can self-reported dieting and dietary restraint identify underreporters of energy intake in dietary surveys? *J Am Diet Assoc* **106**, 1667–1672.
 33. Togo P, Osler M, Sørensen TIA *et al.* (2004) A longitudinal study of food intake patterns and obesity in adult Danish men and women. *Int J Obes Relat Metab Disord* **28**, 583–593.
 34. Drapeau V, Després J-P, Bouchard C *et al.* (2004) Modifications in food-group consumption are related to long-term body-weight changes. *Am J Clin Nutr* **80**, 29–37.
 35. Mamun AA, Kinarivala M, O'Callaghan MJ *et al.* (2010) Associations of excess weight gain during pregnancy with long-term maternal overweight and obesity: evidence from 21 y postpartum follow-up. *Am J Clin Nutr* **91**, 1336–1341.