Published by Oxford University Press on behalf of the International Epidemiological Association International Journal of Epidemiology 2010;**39**:889–899
 Advance Access publication 15 March 2010
doi:10.1093/ije/dyq008

DIABETES

Dietary patterns are associated with lower incidence of type 2 diabetes in middle-aged women: the Shanghai Women's Health Study

Raquel Villegas,[1]* Gong Yang,[1] Yu-Tang Gao,[2] Hui Cai,[1] Honglan Li,[2] Wei Zheng[1] and Xiao Ou Shu[1]

[1]Vanderbilt Epidemiology Center, Department of Medicine, Vanderbilt University Medical Center and Vanderbilt-Ingram Cancer Center, Nashville, TN, USA and [2]Department of Epidemiology, Shanghai Cancer Institute, Shanghai, PR China

*Corresponding author. Vanderbilt Epidemiology Center, 2525 West End Avenue, Suite 600, IMPH, Nashville, TN 37203-1738, USA. E-mail: raquel.villegas@vanderbilt.edu

Accepted 11 January 2010

Background Data linking risk of type 2 diabetes (T2D) and dietary patterns in Chinese populations are scarce.

Methods A population-based prospective study of 64 191 middle-aged women in urban Shanghai, China, who were free of T2D and other chronic diseases at study recruitment, was conducted. Dietary intake, physical activity and anthropometric measurements were assessed through in-person interviews. Dietary patterns were assessed by using K-means cluster analysis. Cox regression model was used to evaluate the association of dietary patterns with the risk of T2D.

Results We identified three dietary clusters in this population. Cluster 1 (56.3%; $N = 36\,159$) had the highest intake of staples, cluster 2 (40.4%: $N = 25\,948$) had the highest intake of dairy milk, and cluster 3 (2.9%; $N = 1843$) had the highest energy intake. Participants in cluster 2 had lower prevalence of obesity, central obesity and hypertension at baseline. Using cluster 1 as the reference, participants in cluster 2 had a lower incidence of T2D after 6.9 years of follow-up [relative risk (RR) 0.78; 95% confidence interval (CI) 0.71–0.86]. The RR for the incidence of T2D for cluster 3 compared with cluster 1 was 1.05 (95% CI 0.81–1.35). The association was not modified by age category, body mass index category, waist-to-hip ratio category or exercise participation.

Conclusions We identified and characterized dietary patterns in middle-aged Chinese women by using cluster analysis. We identified a dietary pattern low in staple foods and high in dairy milk, which was associated with lower risk of T2D. Study of dietary patterns will help elucidate links between diet and disease, and contribute to the development of healthy eating guidelines for health promotion.

Keywords Dietary patterns, cluster analysis, middle-aged women, incidence, type 2 diabetes

Introduction

Although it is well accepted that nutritional factors play an important role in the aetiology of type 2 diabetes (T2D), information about the role of specific dietary compounds is limited. One reason for this is the fact that foods are consumed in combination and, therefore, a particular nutritional factor may not increase the risk of T2D in isolation, but as part of a dietary pattern. The study of dietary patterns addresses the effect of diet as a whole and thus may provide insight beyond the effects described for single nutrients and foods.

Quantitative approaches that have been used to derive dietary patterns include principal component analysis,[1,2] factor analysis,[3–5] cluster analysis[6–11] and reduced rank regression.[12–14] Cluster analysis has an advantage over other quantitative approaches as it aims to identify distinct and relatively homogenous groups based on selected attributes (dietary variables).[15] A Western dietary pattern, characterized by high consumption of high-fat foods, red meat, processed meat and refined grains, as identified by factor analysis, has been associated with a higher incidence of T2D in some studies conducted in the USA.[16,17] Information on dietary patterns and the risk of T2D in Chinese populations is scarce.

The aim of this study was to identify and characterize dietary patterns in a population-based sample of middle-aged women living in urban Shanghai, China, by using cluster analysis. We also report associations of dietary patterns with the incidence of T2D after 6.9 years of follow up and the combined effect of dietary patterns with other low-risk factors for T2D and the incidence of T2D in this population.

Methods

Study population

The Shanghai Women's Health Study (SWHS) is a population-based prospective cohort study of middle-aged women (40–70 years) conducted in seven urban communities in Shanghai, China. Details of the SWHS survey have been reported elsewhere.[18] From a total of 81 170 women who were invited to participate, 75 221 provided written informed consent and enrolled in the study (92.7% participation rate). Reasons for non-participation were refusal (3.0%), absence during the enrolment period (2.6%) and other reasons (health, hearing, speaking problems; 1.6%). After exclusion of women <40 years or >70 years at the time of interview ($N = 278$), 74 942 women remained for the study. Participants completed a detailed survey including an in-person interview for assessment of dietary intake, physical activity and measurement of anthropometrics and other lifestyle factors. Protocols for SWHS were approved by the institutional review boards of all institutes involved in the study. Three biennial in-person follow-ups for all living cohort members were conducted by in-home visits between 2000 and 2002, 2002 and 2004, and 2004 and 2006 with response rates of 99.8, 98.7 and 94.9%, respectively.

Outcome ascertainment

Incident T2D was identified through outcome follow-up surveys. A total of 2273 study participants reported having a T2D diagnosis since the baseline survey and, of those, 2270 participants had valid dietary data. We considered a case of T2D to be confirmed if the participants reported having been diagnosed with T2D and met at least one of the following criteria as recommended by American Diabetes Association[15]: fasting glucose level ⩾7 mmol/l on at least two separate occasions, or an oral glucose tolerance test (OGTT) with a value ⩾11.1 mmol/l, and/or use of hypoglycaemic medication (i.e. insulin or oral hypoglycaemic drugs). Of the self-reported cases a total of 1514 participants met the study outcome criteria and are referred to as 'confirmed' cases of T2D in the present study. Participants from whom information on fasting glucose and OGTT was only available at the second and third follow-up surveys are referred as 'probable' diabetes cases. Because information on the number of abnormal fasting glucose tests and OGTT was not collected in the first follow-up survey, nearly one-third of self-reported cases did not meet our confirmation criteria. Therefore, cases identified during the first follow-up survey could not be confirmed. We performed analyses for all T2D cases and confirmed cases only and found similar results. Thus, in this article we report results with all cases of T2D.

Dietary data

Dietary intake was assessed twice, first during the baseline survey and then at the first follow-up survey via an interviewer-administered food frequency questionnaire (FFQ). The FFQ includes 77 food items and food groups that cover 90% of foods commonly consumed in urban Shanghai during the study period. The FFQ was designed for and validated in this population.[19] To evaluate the validity and reliability of the FFQ, a total of 200 SWHS participants were recruited for a dietary calibration study. Study participants completed an FFQ at baseline and 24-h dietary recalls (24-HDR) twice per month consecutively for 12 months. Validity of the FFQ was evaluated by comparing intake levels of major nutrients and foods obtained from the second FFQ with those derived from the multiple 24-HDR. The median intakes for major nutrients, rice, poultry and meat derived from the second FFQ, and the 24-HDR were similar, with the differences ranging from 1.3 to 12.1%. Nutrient and food intake assessed by the FFQ and the multiple 24-HDR correlated very well, with correlation coefficients of 0.59–0.66 for macronutrients, 0.41–0.59 for micronutrients and 0.41–0.66 for major food groups. The reliability of the FFQ was assessed by comparing the correlation and median intake of nutrients and food groups obtained from the two FFQs that were administered ~2 years apart. The median intake levels for selected nutrients and food groups derived from the two FFQs were similar, with differences <10%. When nutrient and food-group intakes were categorized into quartiles, FFQ and 24-HDR produced exact agreement rates between 33 and 50%. Misclassification to an adjacent quartile was common, ranging from 34 to 48%, whereas misclassification to an extreme quartile was rare (1–6%).

If women had been diagnosed with T2D, cancer or cardiovascular disease between baseline and the follow-up surveys, we used dietary data from the baseline FFQ in the analysis. For other participants we used the average of baseline and follow-up FFQ data. The Chinese Food Composition Tables[20] were used to estimate intake of energy (kcal/day) and other nutrients. We excluded from this study participants who had extreme values for total energy intake (<500 or >3500 kcal/day; $N = 36$),[21] leaving 64 191 participants for the final analysis. Items from the FFQ as expressed in terms of proportion of total mass of food consumed (g/day) were aggregated into 11 mutually exclusive food groups: staples (rice, noodles, steamed bread, bread), pork, red meat (other than pork), poultry, eggs, fish and shellfish, soy foods (soy beans, bean curd, soy milk), vegetables, fruits, dairy milk and snacks/desserts (peanuts, candy, preserved fruit and desserts).

Other factors as potential confounders

Anthropometric measurements, including weight, height and circumferences of waist and hips, were taken at baseline recruitment according to a standard protocol by trained interviewers who were retired medical professionals.[22] From these measurements, the following variables were created: body mass index (BMI; weight in kilograms divided by the square of height in metres), and waist-to-hip ratio (WHR; waist circumference divided by hip circumference).

Detailed information on physical activity was collected by using a validated questionnaire.[23] The validity of the questionnaire was evaluated by comparing Spearman correlations (r) for the physical activity questionnaire with two criterion measures administered over a period of 12 months (four 7-day physical activity logs and up to 28 7-day physical activity questionnaires). Significant correlations between the physical activity questionnaire and criterion measures for exercise were observed (physical activity log, $r = 0.74$; 7-day physical activity questionnaire, $r = 0.80$). Significant correlations between activities on the physical activity questionnaire lifestyle and activities on the 7-day questionnaire were also observed ($r = 0.30$–0.88). The reproducibility of the questionnaire (2-year test–retest) was evaluated using kappa statistics and intraclass correlation coefficients ($\kappa = 0.64$; intraclass correlations coefficients: 0.14–0.54).[23]

The questionnaire evaluated regular exercise and sports participation during the 5 years preceding the interview, daily activity and daily commuting activity (round-trip journey to and from work). We calculated the metabolic equivalents (METs) for each activity, using a compendium of physical activity values.[24] One MET-h/day is roughly equivalent to 1 kcal/kg/day or ~15 min of participation in moderate intensity (four METs) activity for an average adult.[24] We combined each of the exercise and lifestyle activity indices to derive a quantitative estimate of overall non-occupational physical activity (MET-h/day).

Information on socio-demographic factors and other potential confounders such as age, level of education (none, elementary school, middle/high school, college), family income in yuan/year (<10 000, 10 000–19 999, 20 000–29 999, >30 000), occupation (professional, clerical, manual labourers, housewife/retired), smoking (ever smoked one or more cigarette per day for >6 months continuously), alcohol intake (ever drank beer, wine or spirits three or more times per week) and presence of hypertension at baseline was collected by using a structured questionnaire.

Statistical analysis

From the 64 227 participants who were free of T2D and other chronic diseases (cancer and cardiovascular disease) at baseline, we excluded participants that had extreme values for total energy intake (<500 or >3500 kcal/day; $N = 36$),[16] leaving 64 191 participants for the analysis.

We used cluster analysis to identify dietary patterns and to segregate participants based on similarity of diet. Continuous food groups were standardized by converting to the standard normal deviate to ensure that clusters were not influenced by food groups with high specific gravity. We chose food variables because we wanted to identify food patterns clusters. K-means cluster analysis was used to define clusters of participants by using the cluster analysis procedure in Statistical Software (SAS). This procedure attempts to identify relatively homogeneous groups of participants based on selected characteristics, using an algorithm that can handle large numbers of cases. In K-means cluster analysis the homogeneity of cases within a cluster is measured by the total within-cluster sum of squares. Cluster memberships are determined by sequentially moving cases from one cluster to another so that the total within-cluster sum of squares is minimised. The algorithm requires the number of clusters to be specified prior to analysis.

To identify the optimal number of clusters, we conducted several runs with varying number of clusters. Each set of clusters was examined to find a reasonably sized solution. A cluster solution of four clusters, derived three major clusters and one small cluster ($N = 241$, 0.38%). The small cluster was not included in subsequent analyses. To evaluate the robustness of the clusters, we performed a cluster analysis with a cluster solution of four clusters in a random sample of 50% of participants and derived clusters similar to those in the original analysis. We repeated this procedure a few times. Similar clusters to those in the original analysis were found 80% of the time.

Differences in food-group consumption were investigated using a non-parametric analysis of variance, the Kruskal–Wallis test. We investigated differences between clusters among continuous variables by using analysis of variance or the Kruskal–Wallis test, if the data were not normally distributed. We used the chi-square test to assess differences between clusters in categorical lifestyle factors and socio-economic status variables.

Person-years of follow-up for each participant were calculated as the interval between the baseline recruitment to the diagnosis of T2D, censored at death or completion of the third follow-up. The Cox proportional hazards model was used to assess the effect of dietary patterns on the incidence of T2D. We adjusted the analyses for socio-demographic factors and known T2D risk factors as potential confounders.

We conducted analyses stratified by age group, BMI, WHR and exercise participation. The log-likelihood ratio test was used to evaluate the significance of these interaction terms. We also evaluated the combined effect of factors associated with a low risk of T2D. Factors included in the analysis were normal weight (cut-off point 23 for Asian populations[25]), low WHR (cut-off point 0.85[26]), exercise participation and a dietary cluster associated with lower T2D risk in this population. All analyses were performed using SAS (version 9.1), and all tests of statistical significance were based on two-sided probability.

Results

K-cluster analysis identified three distinct groups in this population. A total of 36 159 participants (56.3%) were in cluster 1, 25 948 (40.4%) in cluster 2 and 1843 (2.9%) in cluster 3. Medians for consumption (g/day) of each food group in each cluster are presented in Table 1. The Kruskal–Wallis test revealed that food groups were consumed at different levels across the three clusters (data not shown). Cluster 1 had the highest median intake of staples, cluster 2 had the highest intake of dairy milk and cluster 3 had the highest intake of poultry, pork, red meat, fish and shellfish, eggs, fruits and vegetables and the highest energy intake. Nutrient intakes also varied by cluster, with cluster 1 having the lowest intake of vitamins and minerals and the highest carbohydrate intake (Table 1).

Cluster 2 had the lowest prevalence of participants with general or central obesity and hypertension (Table 2). There were fewer smokers in cluster 2, whereas cluster 3 had the highest proportion of alcohol drinkers. We also found differences in socio-economic status characteristics among clusters. Participants in cluster 2 had higher income and were more likely to hold professional jobs and have a college education, whereas participants in cluster 1 had less education, lower income and were more likely to be retired or not employed at the time of the survey. Nutrient intake varied significantly across the clusters. Participants in cluster 1 had the lowest intakes of fibre, energy from fat, energy from protein, minerals and vitamins, and the highest percentage of energy intake from carbohydrates. Participants in cluster 1 were also less likely to take vitamin supplements. Participants in cluster 3 had the highest intake of all nutrients with the exception of carbohydrates. Cluster 3 participants reported a higher consumption of processed meats and were more likely to roast or fry their foods as compared with clusters 1 and 2 (data not shown).

When cluster 1 was used as the reference group, participants in cluster 2 had a lower incidence of T2D (Table 3), whereas no association between the risk of T2D and being in cluster 3 was found. In fully adjusted analyses, including adjustments for BMI and WHR, the relative risks for clusters 2 and 3 were 0.78 [95% confidence interval (CI) 0.71–0.86] and 1.05 (95% CI 0.81–1.35), respectively. We stratified analyses by age group (<50 and $\geqslant 50$ years) and found similar results. Sensitivity analyses, in which we excluded participants who had been diagnosed with T2D in the first 2 years of follow-up, showed similar results (data not shown in tables). In participants with low BMI ($\leqslant 23$), cluster 2 was not associated with a lower risk of T2D when compared with cluster 1, although the coefficient of interaction was not significant (Table 4). The association between dietary patterns and T2D did not differ by exercise participation or WHR categories.

When we investigated the combined effect of having a 'protective dietary pattern' for T2D with other factors associated with a lower risk for T2D (BMI $\leqslant 23$, WHR < 0.85, and exercise participation), we found an inverse dose–response relationship between the number of factors associated with lower risk of T2D and T2D incidence (Table 5). The relative risk (RR) for having none, 1, 2, 3 and 4 low-risk T2D factors w 1.00, 0.66, 0.38, 0.22 and 0.14, respectively (P trend < 0.001).

Discussion

In this prospective study we identified three dietary patterns by cluster analysis. The first dietary pattern had a higher intake of staple foods, the second dietary pattern had a higher intake of dairy milk and the

Table 1 Food-group intake,[a] nutrient intake,[b] total daily energy intake[b] and percentage of vitamin supplement use by dietary clusters

	Cluster 1 $N = 36\,159$	Cluster 2 $N = 25\,948$	Cluster 3 $N = 1843$
Staple foods (g/day)	303.6	266.4	286.2
Rice (g/day)	250.0	225.0	225.0
Noodles (g/day)	26.6	25.0	32.1
Bread (g/day)	4.6	11.5	14.0
Poultry (g/day)	9.3	14.4	65.5
Pork (g/day)	37.3	40.3	63.0
Red meat[c] (g/day)	2.0	3.3	8.3
Fish and shellfish (g/day)	33.1	53.0	91.0
Fish (g/day)	25.6	39.8	69.1
Shellfish (g/day)	5.6	10.8	14.3
Eggs (g/day)	23.3	31.2	38.4
Fruits (g/day)	192.2	282.7	343.0
Vegetables (g/day)	249.6	277.3	384.7
Soy food(g/day)	146.9	80.8	144.3
Soy beans (g/day)	9.7	10.2	14.0
Bean curd (g/day)	48.9	42.9	69.1
Soy milk (g/day)	71.4	10.2	49.1
Dairy milk (g/day)	12.3	129.1	108.2
Snacks and desserts (g/day)	9.5	24.2	22.3
Carbohydrates (% energy)	69.9 ± 0.03	65.7 ± 0.03	58.6 ± 0.13
Protein (% energy)	15.5 ± 0.01	16.7 ± 0.01	19.6 ± 0.05
Fat (% energy)	14.5 ± 0.02	17.6 ± 0.02	21.7 ± 0.09
Fiber (g/day)	10.4 ± 0.02	11.2 ± 0.02	14.5 ± 0.1
Carotene (μg/day)	2730.7 ± 0.2	3077.2 ± 8.2	4034.6 ± 41.7
Vitamin C (mg/day)	81.6 ± 0.2	98.0 ± 0.3	131.0 ± 1.3
Vitamin E (mg/day)	13.0 ± 0.03	13.9 ± 0.03	19.0 ± 0.1
Magnesium (mg/day)	265.0 ± 0.3	281.3 ± 0.5	368.2 ± 2.0
Calcium (mg/day)	432.5 ± 0.9	537.4 ± 1.0	677.5 ± 5.0
Daily energy intake (kcal/day)	1623 ± 1.8	1660 ± 2.1	2062 ± 9.0

[a]Median.
[b]Mean ± standard error.
[c]Other than pork.

third dietary pattern had a higher intake of most other food groups. The dietary patterns were associated with distinct nutrient intake profiles and differences in socio-economic status, overall obesity, central obesity and hypertension prevalence. The cluster 2 dietary pattern was associated with a lower risk of diabetes after 6.9 years of follow-up. We found that combining this dietary pattern with other known low-risk factors ($\text{BMI} \leqslant 23$, $\text{WHR} < 0.85$ and exercise participation) resulted in an 86% reduction in risk of T2D.

Our results are in agreement with previous studies reporting that dietary patterns were related to other behaviours or socio-demographic characteristics.[2,15,27] Participants in cluster 2 were less likely to have ever smoked or to have general or central obesity than those in other dietary clusters, whereas participants in cluster 3 (the smallest cluster) were more likely to drink alcohol. We did not find significant differences in levels of physical activity among the clusters. Participants in cluster 1, the cluster with highest intake of staples and soy foods, had characteristics reflecting lower socio-economic status (more likely to have lower income, less likely to have a professional job and less likely to have lower education level). It has been suggested that a high intake of refined grains in China is more common in rural populations and lower-income segments of the population.[28]

Table 2 Characteristics of study participants by dietary clusters

	Cluster 1	Cluster 2	Cluster 3	P^a
Age (years, mean ± SE)	52.5 ± 0.05	49.1 ± 0.05	48.1 ± 0.17	<0.001
Education (%)				<0.001
None	13.6	2.9	2.1	
Elementary school	12.8	5.2	4.1	
Up to high school	63.6	73.1	76.5	
College	9.9	18.8	17.4	
Income level (%)				
<10 000	18.5	11.0	13.9	<0.001
10 000–19 999	40.9	34.4	33.2	
20 000–29 999	26.4	31.7	31.4	
>30 000	14.2	22.9	21.4	
Occupation (%)				
Professional	15.0	26.4	23.6	<0.001
Clerical	11.8	14.5	15.1	
Manual	21.9	24.1	26.8	
Housewife/retired	51.2	35.0	35.0	
Smoking (%)	2.7	1.5	2.2	<0.001
Alcohol consumption (%)	2.1	2.3	4.6	<0.001
Exercise participation (%)	32.9	33.0	32.4	0.84
Obesity (%)[b]	38.1	25.2	29.7	<0.001
Central obesity (%)[c]	23.4	14.0	17.2	<0.001
Hypertension (%)	22.1	14.56	15.7	<0.001

SE: standard error.
[a]*P*-value was calculated by ANOVA or the Kruskal–Wallis test (for non-normally distributed variables) for continuous variables and the chi-square test for categorical variables.
[b]BMI > 25 kg/m^2.
[c]WHR ⩾ 0.85.

Table 3 Associations between dietary clusters and the incidence of T2D stratified by age group[a]

	Person-years	Cases	RR1[b]	95% CI	RR2[c]	95% CI
All						
Cluster 1	258 575	1588	1.00		1.00	
Cluster 2	190 078	612	0.68	0.62–0.75	0.78	0.71–0.86
Cluster 3	13 302	67	0.95	0.74–1.22	1.05	0.81–1.35
Age < 50 years						
Cluster 1	121 937	427	1.00		1.00	
Cluster 2	120 448	234	0.64	0.54–0.75	0.77	0.66–0.91
Cluster 3	9184	27	0.82	0.55–1.23	0.98	0.66–1.47
Age ⩾ 50 years						
Cluster 1	136 638	1161	1.00		1.00	
Cluster 2	69 630	378	0.74	0.66–0.84	0.83	0.73–0.93
Cluster 3	4118	40	1.15	0.83–1.59	1.19	0.86–1.65

[a]*P* interaction 0.10.
[b]RR1 adjusted for age, kcal/day, physical activity, alcohol consumption, smoking, education level, income level, occupation and hypertension.
[c]RR2 as above plus adjustment for WHR and BMI.

Several studies have reported on dietary patterns in Western populations. In general, two main patterns have been reported, including a prudent or healthy dietary pattern, characterised by higher intake of fruits and vegetables and lower intake of fried or high-fat foods,[2,11,14,15,27] as opposed to a Western, traditional or conservative pattern, which is characterized by higher intake of fried foods, red meat and sweets/desserts.[4,29,30] The prudent dietary pattern has been associated with a lower risk of T2D,[2,14,17,29] whereas the Western diet has been associated with a higher risk of T2D.[16,17] Patterns identified by reduced rank regression (RRR) that were similar to the Western pattern have been associated with a higher risk of inflammation and insulin resistance.[31–33]

Table 4 Relative risk of diabetes by clusters stratified by BMI categories, WHR categories and exercise participation categories

	RR	95% CI	RR	95% CI
	BMI ⩽ 23		BMI >23	
Cluster 1	1.00		1.00	
Cluster 2	0.88	0.70–1.11	0.73	0.66–0.82
Cluster 3	1.19	0.63–2.22	1.00	0.76–1.32
P interaction = 0.95				
	WHR < 0.85		WHR ⩾ 0.85	
Cluster 1	1.00		1.00	
Cluster 2	0.76	0.67–0.87	0.84	0.72–0.98
Cluster 3	1.20	0.88–1.64	0.84	0.57–1.30
P interaction = 0.20				
	No exercise		Exercise	
Cluster 1	1.00		1.00	
Cluster 2	0.78	0.69–0.88	0.82	0.70–0.95
Cluster 3	1.01	0.735–1.39	1.13	0.75–1.71
P interaction = 0.81				

Analyses adjusted for age, kcal/day, alcohol consumption, smoking, education level, income level, occupation and hypertension. In addition, the analyses stratified by BMI and WHR were adjusted for physical activity and for each other; the analysis stratified by physical activity was also adjusted for BMI and WHR.

Data describing dietary patterns in Asian populations are scarce. Some data are available on dietary patterns in Japan[34–39] and China, including our population.[28,40,41] Using baseline dietary information and principal component analysis (PCA), we previously reported three major dietary patterns in this population: vegetable-rich, fruit-rich and meat-rich.[41] In the present study, we derived dietary patterns by using cluster analysis so that we could identify distinct groups and evaluate their associations with incidence of T2D. This also allowed us to estimate the combined effect of particular dietary patterns with other protective risk factors in the development of T2D. One small study of 119 men and women living in Beijing that also used cluster analysis derived three dietary patterns: 'fruit and milk', 'red meat' and 'refined cereal'. In that study, the 'refined cereal' cluster was characterized by a high intake of white rice, noodles, and buns and bread, which were the food items included in staples food group in our study.[28]

We found cluster 2 to be associated with a lower risk of T2D when compared with cluster 1 in this population. One reason for this could be related to the difference in the intake of staples in these dietary patterns. The median intake of staples for cluster 1 was 303.6 g/day and for cluster 2 was 266.0 g/day, and rice intake was 225 g/day in cluster 2 as opposed to 250 g/day in cluster 1. We have previously reported in this population that the intake of staples, and rice in particular, was associated with a higher risk of T2D.[42] When we investigated the association between the three principal components previously indentified in this population[41] and the incidence of T2D, we found an inverse association between two of the components and T2D incidence. The RR of diabetes for the second, third, fourth and fifth quintiles compared with the first quintile were 1.00, 1.03, 1.04, 1.06 and 1.05, $P = 0.37$ for the vegetable-rich pattern; 1.00, 0.82, 0.88, 0.87 and 0.81, $P = 0.01$ for the fruit-rich pattern; and 1.00, 0.98, 0.87, 0.83 and 0.89, $P < 0.01$ for the meat-rich pattern.[41] The factor loading for rice in these two components (the fruit-rich and meat-rich patterns) was –0.40, which also suggests that a lower intake of rice might be associated with a lower risk of T2D in this population. Similarly, a recent study of dietary patterns in Japan reported that a dietary

Table 5 RR of diabetes by number of low-risk factors[a]

	Person-years	Cases	RR[b]	95% CI	P trend
None	32 554	435	1.00		0.001
1 low-risk factor	113 343	880	0.66	0.59–0.74	
2 low-risk factors	169 639	665	0.38	0.33–0.42	
3 low-risk factors	118 881	251	0.22	0.19–0.26	
4 low-risk factors	27 538	36	0.14	0.10–0.19	

[a]Low risk factors: BMI ⩽ 23, low WHR, exercise participation, and member of cluster 2.

[b]RR adjusted for age, kcal/day, alcohol consumption, smoking, education level, income level, occupation and hypertension.

pattern that was characterized by lower intake of rice was associated with lower levels of glycated haemoglobin.[35] Data from a Japanese population reported an inverse association with glucose intolerance for a dietary pattern with high intakes of dairy, fruits, vegetables and starch, and low intake of alcohol (*P* for trend < 0.0001). The odds ratios (ORs) for having a glucose tolerance abnormality (impaired fasting glucose, impaired glucose tolerance or T2D) for the second, third and fourth quartiles were 0.80 (95% CI 0.62–1.04), 0.71 (95% CI 0.54–0.92) and 0.51 (95% CI 0.38–0.67), respectively, compared with the lowest quartile. A 'Japanese dietary pattern' was positively associated with impaired glucose tolerance.[39] The authors indicated that a Japanese diet might be characterized by being high in refined carbohydrates and low in protein, which is similar to cluster 1 in our study. In the study that identified dietary patterns using cluster analysis in Beijing, the refined cereals pattern was associated with higher homocysteine levels, a marker of risk for cardiovascular disease.[28]

Cluster 1 also had a higher intake of soy foods than cluster 2. Although soy foods contain beneficial nutrients, such as phytoestrogens, these could be negatively impacted by consumption of food groups such as refined grains due to nutrient displacement (e.g. fibre). Thus, it is possible that the detrimental effect of a diet with high consumption of staple foods outweighs any protective effect or beneficial components from soy foods. However, in a previous report, soy beans and soy milk were associated with a lower incidence of T2D after 4.6 years of follow-up. No association between the intake of soy products (bean curd) or soy protein (reflecting total soy intake) with T2D was found in this population.[43]

When compared with cluster 1, cluster 2 also had a higher intake of other food groups, such as vegetables and dairy milk, which have been associated with a lower risk of diabetes in our study population. Milk is a good source of magnesium and calcium, and those two nutrients have been associated with lower risk of T2D in some studies,[44–47] including in our population.[48] Vegetable intake was also associated with a lower risk of T2D in our this population.[49] Several studies have shown that vegetables have a high content of antioxidants,[50] fibre[51] and magnesium,[52] all of which reduce the risk of T2D. In our study, intake of those nutrients was higher in cluster 2 compared with cluster 1.

Cluster 2 also had a higher intake of snacks/desserts (including peanuts, desserts, candy and preserved fruit) than cluster 1. We found peanuts to be associated with a lower risk of T2D[43] in this study, but did not find an association between sweets or desserts (unpublished data). Desserts and sweets have been found to be part of the Western dietary pattern associated with higher T2D risk.[16,17] However, in our population consumption of desserts is low (median 10.7 g/day) compared with Western populations.

Although in Shanghai dietary patterns are different from those in Western populations, the smallest cluster (cluster 3) had a higher proportion of alcohol drinkers, higher consumption of processed foods and higher daily energy intake. This cluster was associated with a higher risk of T2D when compared with cluster 2 (data not shown in tables). We have previously reported that processed meat consumption was associated with a higher risk of T2D in this population.[53] This group of women was also more likely to fry their foods than those in the other two clusters. Consumption of fried foods and processed foods have been associated with an overall dietary pattern linked to adverse health outcomes, including diabetes.[2,16,17]

Our results on the combined effect of low-risk factors are consistent with findings from the Nurses Health Study, where a 90% reduction in risk of T2D was observed in women with five protective factors compared with all other participants after 16 years of follow-up.[54] Intervention trials have also reported benefits associated with the combined effects of diet and physical activity in prevention of T2D in high risk populations.[55–58]

It should be noted that the reproducibility and stability of dietary patterns derived by cluster analysis is not clear,[59] whereas dietary patterns derived using PCA have reasonable reproducibility and validity. We chose to use cluster analysis for this study, because this approach identifies relatively homogeneous groups based on individual differences in mean intake and the results are easier to interpret than results derived from PCA. PCA is a data reduction technique that reduces the complexity of diet to a small number of dimensions, based on the inter-correlations between dietary items.

Strengths of our study include the following: our participants are representative of the middle-aged, female, Chinese population; the participation rate and follow-up rates were very high; we measured potentially important confounders; and we used repeated dietary measurements.

Limitations of our analysis should also be noted. A major concern is reliance on self-reports of T2D. In addition, imperfections in dietary assessment are always a concern in observational epidemiology. Although these imperfections are generally thought to be randomly distributed among categories of the outcome (non-differential), systematic imperfections that could bias risk estimates toward or away from the null value are a possibility. In addition, there is concern about the possibility of residual confounding. Because of the strong associations between the dietary clusters and other health behaviours in our population, it is likely that statistical adjustment cannot adequately control for confounding. Thus, risk estimates may be confounded by demographic and non-dietary lifestyle factors and might have been inadequately captured with the covariates we included in our multivariable models. There might also be limitations associated with the method used to derive the dietary patterns, and it should be noted that the reproducibility and stability of dietary patterns derived by cluster analysis are not clear.[59]

In conclusion, we have shown that cluster analysis is effective in identifying major dietary patterns that segregate individuals within our population into groups with significant differences in intake of nutrients, smoking, alcohol intake, obesity prevalence and socio-economic factors. We identified a dietary pattern that was associated with a lower risk of T2D in this population. A favourable dietary pattern combined with low BMI, low WHR and exercise participation was associated with an 86% reduction of T2D in this population.

The principle implications of our findings highlight the importance of dietary patterns in the prevention of T2D in China and the combined effects of risk factors and the risk of T2D. The current study addresses an important issue, since Chinese people are facing a rapid nutritional and lifestyle transition accompanied by an increase in obesity.[60] As a consequence, T2D has become a major public health issue in China. The prevalence of T2D increased from 1.9 to 5.6% between 1993 and 2003 in China.[61] Thus, prevention of T2D through lifestyle factors should be a high public health priority.

Funding

Grant number R01 CA070867 from the National Cancer Institute and R01 HL079123 from the National Heart, Lung and Blood Institute.

Acknowledgements

The authors thank the participants and research staff of the Shanghai Women's Health Study for their contributions to the study and Ms Bethanie Hull for technical assistance in the preparation of this manuscript. They also thank Dr FB Hu for his advice on this work.

Conflict of interest: None declared.

KEY MESSAGES

- Chinese people are facing a rapid nutritional and lifestyle transition accompanied by an increase in obesity and T2D has become a major public health issue in China.
- A dietary pattern low in staples and high in dairy milk in middle age Chinese women was associated with a lower risk of T2D in this population.
- This dietary pattern combined with low BMI, low WHR, and exercise participation was associated with an 86% reduction of T2D in this population.
- Prevention of type 2 diabetes might be possible by adopting a number of 'protective' factors.

References

1 Barker ME, McClean SI, Strain JJ, Thompson KA. Dietary behaviour and health in Northern Ireland: an exploration of biochemical and haematological associations. *J Epidemiol Community Health* 1992;**46:**151–56.

2 Williams DE, Prevost AT, Whichelow MJ, Cox BD, Day NE, Wareham NJ. A cross-sectional study of dietary patterns with glucose intolerance and other features of the metabolic syndrome. *Br J Nutr* 2000;**83:** 257–66.

3 Fung TT, Rimm EB, Spiegelman D *et al*. Association between dietary patterns and plasma biomarkers of obesity and cardiovascular disease risk. *Am J Clin Nutr* 2001; **73:**61–67.

4 Hu FB, Rimm EB, Stampfer MJ, Ascherio A, Spiegelman D, Willett WC. Prospective study of major dietary patterns and risk of coronary heart disease in men. *Am J Clin Nutr* 2000;**72:**912–21.

5 Randall E, Marshall JR, Graham S, Brasure J. Patterns in food use and their associations with nutrient intakes. *Am J Clin Nutr* 1990;**52:**739–45.

6 Akin JS, Guilkey DK, Popkin BM, Fanelli MT. Cluster analysis of food consumption patterns of older Americans. *J Am Diet Assoc* 1986;**86:**616–24.

7 Farchi G, Mariotti S, Menotti A, Seccareccia F, Torsello S, Fidanza F. Diet and 20-y mortality in two rural population groups of middle-aged men in Italy. *Am J Clin Nutr* 1989;**50:**1095–103.

8 Huijbregts PP, Feskens EJ, Kromhout D. Dietary patterns and cardiovascular risk factors in elderly men: the Zutphen Elderly Study. *Int J Epidemiol* 1995;**24:**313–20.

9 Hulshof KF, Wedel M, Lowik MR *et al*. Clustering of dietary variables and other lifestyle factors (Dutch Nutritional Surveillance System). *J Epidemiol Community Health* 1992; **46:**417–24.

10 Millen BE, Quatromoni PA, Gagnon DR, Cupples LA, Franz MM, D'Agostino RB. Dietary patterns of men and women suggest targets for health promotion: the Framingham Nutrition Studies. *Am J Health Promot* 1996;**11:**42–52.

11 Pryer JA, Nichols R, Elliott P, Thakrar B, Brunner E, Marmot M. Dietary patterns among a national random

sample of British adults. *J Epidemiol Community Health* 2001;**55:**29–37.

12 Heidemann C, Hoffmann K, Spranger J *et al*. A dietary pattern protective against type 2 diabetes in the European Prospective Investigation into Cancer and Nutrition (EPIC)—Potsdam Study cohort. *Diabetologia* 2005;**48:** 1126–34.

13 Hoffmann K, Schulze MB, Schienkiewitz A, Nothlings U, Boeing H. Application of a new statistical method to derive dietary patterns in nutritional epidemiology. *Am J Epidemiol* 2004;**159:**935–44.

14 Nettleton JA, Steffen LM, Ni H, Liu K, Jacobs DR Jr. Dietary patterns and risk of incident type 2 diabetes in the Multi-Ethnic Study of Atherosclerosis (MESA). *Diabetes Care* 2008;**31:**1777–82.

15 Tucker KL, Dallal GE, Rush D. Dietary patterns of elderly Boston-area residents defined by cluster analysis. *J Am Diet Assoc* 1992;**92:**1487–91.

16 Fung TT, Schulze M, Manson JE, Willett WC, Hu FB. Dietary patterns, meat intake, and the risk of type 2 diabetes in women. *Arch Intern Med* 2004;**164:**2235–40.

17 van Dam RM, Rimm EB, Willett WC, Stampfer MJ, Hu FB. Dietary patterns and risk for type 2 diabetes mellitus in U.S. men. *Ann Intern Med* 2002;**136:**201–9.

18 Zheng W, Chow WH, Yang G *et al*. The Shanghai Women's Health Study: rationale, study design, and baseline characteristics. *Am J Epidemiol* 2005;**162:**1123–31.

19 Shu XO, Yang G, Jin F *et al*. Validity and reproducibility of the food frequency questionnaire used in the Shanghai Women's Health Study. *Eur J Clin Nutr* 2004; **58:**17–23.

20 Yang YX, Wang GY, Pan XC. *China Food Composition*. Beijing: Beijing University Medical Press, 2002.

21 Joshipura KJ, Hu FB, Manson JE *et al*. The effect of fruit and vegetable intake on risk for coronary heart disease. *Ann Intern Med* 2001;**134:**1106–14.

22 Zhang X, Shu XO, Gao YT *et al*. Anthropometric predictors of coronary heart disease in Chinese women. *Int J Obes Relat Metab Disord* 2004;**28:**734–40.

23 Matthews CE, Shu XO, Yang G *et al*. Reproducibility and validity of the Shanghai Women's Health Study physical activity questionnaire. *Am J Epidemiol* 2003;**158:** 1114–22.

24 Ainsworth BE, Haskell WL, Whitt MC *et al*. Compendium of physical activities: an update of activity codes and MET intensities. *Med Sci Sports Exerc* 2000;**32:**S498–S504.

25 WHO Expert Consultation. Appropriate body-mass index for Asian populations and its implications for policy and intervention strategies. *Lancet* 2004;**363**:157–63.

26 WHO. *Definition, Diagnosis and Classification of Diabetes Mellitus and Its Complications*. Geneva: World Health Organisation, 1999.

27 Margetts BM, Thompson RL, Speller V, McVey D. Factors which influence 'healthy' eating patterns: results from the 1993 Health Education Authority health and lifestyle survey in England. *Public Health Nutr* 1998;**1:**193–98.

28 Gao X, Yao M, McCrory MA *et al*. Dietary pattern is associated with homocysteine and B vitamin status in an urban Chinese population. *J Nutr* 2003;**133:**3636–42.

29 Montonen J, Knekt P, Harkanen T *et al*. Dietary patterns and the incidence of type 2 diabetes. *Am J Epidemiol* 2005; **161:**219–27.

30 Slattery ML, Boucher KM, Caan BJ, Potter JD, Ma KN. Eating patterns and risk of colon cancer. *Am J Epidemiol* 1998;**148:**4–16.

31 McNaughton SA, Mishra GD, Brunner EJ. Dietary patterns, insulin resistance and incidence of type 2 diabetes in the Whitehall II study. *Diabetes Care* 2008.

32 Nettleton JA, Steffen LM, Schulze MB *et al*. Associations between markers of subclinical atherosclerosis and dietary patterns derived by principal components analysis and reduced rank regression in the Multi-Ethnic Study of Atherosclerosis (MESA). *Am J Clin Nutr* 2007;**85:** 1615–25.

33 Schulze MB, Hoffmann K, Manson JE *et al*. Dietary pattern, inflammation, and incidence of type 2 diabetes in women. *Am J Clin Nutr* 2005;**82:**675–84.

34 Shimazu T, Kuriyama S, Hozawa A *et al*. Dietary patterns and cardiovascular disease mortality in Japan: a prospective cohort study. *Int J Epidemiol* 2007;**36:**600–9.

35 Nanri A, Mizoue T, Yoshida D, Takahashi R, Takayanagi R. Dietary patterns and glycated hemoglobin in Japanese men and women. *Diabetes Care* 2008;**31:** 1568–73.

36 Nanri A, Yoshida D, Yamaji T, Mizoue T, Takayanagi R, Kono S. Dietary patterns and C-reactive protein in Japanese men and women. *Am J Clin Nutr* 2008;**87:** 1488–96.

37 Sadakane A, Tsutsumi A, Gotoh T *et al*. Dietary patterns and levels of blood pressure and serum lipids in a Japanese population. *J Epidemiol* 2008;**18:**58–67.

38 Okubo H, Sasaki S, Murakami K *et al*. Three major dietary patterns are all independently related to the risk of obesity among 3760 Japanese women aged 18-20 years. *Int J Obes (Lond)* 2008;**32:**541–49.

39 Mizoue T, Yamaji T, Tabata S *et al*. Dietary patterns and glucose tolerance abnormalities in Japanese men. *J Nutr* 2006;**136:**1352–58.

40 Cai H, Zheng W, Xiang YB *et al*. Dietary patterns and their correlates among middle-aged and elderly Chinese men: a report from the Shanghai Men's Health Study. *Br J Nutr* 2007;**98:**1006–13.

41 Cai H, Shu XO, Gao YT, Li H, Yang G, Zheng W. A prospective study of dietary patterns and mortality in Chinese women. *Epidemiology* 2007;**18:**393–401.

42 Villegas R, Liu S, Gao YT *et al*. Prospective study of dietary carbohydrates, glycemic index, glycemic load, and incidence of type 2 diabetes mellitus in middle-aged Chinese women. *Arch Intern Med* 2007;**167:** 2310–6.

43 Villegas R, Gao YT, Yang G *et al*. Legume and soy food intake and the incidence of type 2 diabetes in the Shanghai Women's Health Study. *Am J Clin Nutr* 2008; **87:**162–67.

44 Bo S, Pisu E. Role of dietary magnesium in cardiovascular disease prevention, insulin sensitivity and diabetes. *Curr Opin Lipidol* 2008;**19:**50–56.

45 van Dam RM, Hu FB, Rosenberg L, Krishnan S, Palmer JR. Dietary calcium and magnesium, major food sources, and risk of type 2 diabetes in U.S. black women. *Diabetes Care* 2006;**29:**2238–43.

46 Lopez-Ridaura R, Willett WC, Rimm EB *et al*. Magnesium intake and risk of type 2 diabetes in men and women. *Diabetes Care* 2004;**27:**134–40.

[47] Song Y, Manson JE, Buring JE, Liu S. Dietary magnesium intake in relation to plasma insulin levels and risk of type 2 diabetes in women. *Diabetes Care* 2004;**27:** 59–65.

[48] Villegas GDYCLZS. Dietary calcium and magnesium intakes and the risk of type 2 diabetes: the Shanghai Women's Health Study. *Am J Clin Nutr* 2009; in press.

[49] Villegas R, Shu XO, Gao YT *et al*. Vegetable but not fruit consumption reduces the risk of type 2 diabetes in Chinese women. *J Nutr* 2008;**138:**574–80.

[50] Walker KZ, O'Dea K, Nicholson GC, Muir JG. Dietary composition, body weight, and NIDDM. Comparison of high-fiber, high-carbohydrate, and modified-fat diets. *Diabetes Care* 1995;**18:**401–3.

[51] Hu FB, van Dam RM, Liu S. Diet and risk of Type II diabetes: the role of types of fat and carbohydrate. *Diabetologia* 2001;**44:**805–17.

[52] Lopez-Ridaura R, Willett WC, Rimm EB *et al*. Magnesium intake and risk of type 2 diabetes in men and women. *Diabetes Care* 2004;**27:**134–40.

[53] Villegas R, Shu XO, Gao YT *et al*. The association of meat intake and the risk of type 2 diabetes may be modified by body weight. *Int J Med Sci* 2006;**3:**152–59.

[54] Hu FB, Manson JE, Stampfer MJ *et al*. Diet, lifestyle, and the risk of type 2 diabetes mellitus in women. *N Engl J Med* 2001;**345:**790–97.

[55] Eriksson KF, Lindgarde F. Prevention of type 2 (non-insulin-dependent) diabetes mellitus by diet and physical exercise. The 6-year Malmo feasibility study. *Diabetologia* 1991;**34:**891–98.

[56] Pan X, Li G, Hu Y. Effect of dietary and/or exercise intervention on incidence of diabetes in 530 subjects with impaired glucose tolerance from 1986-1992. *Zhonghua Nei Ke Za Zhi* 1995;**34:**108–12.

[57] Pan XR, Li GW, Hu YH *et al*. Effects of diet and exercise in preventing NIDDM in people with impaired glucose tolerance. The Da Qing IGT and Diabetes Study. *Diabetes Care* 1997;**20:**537–44.

[58] Tuomilehto J, Lindstrom J, Eriksson JG *et al*. Prevention of type 2 diabetes mellitus by changes in lifestyle among subjects with impaired glucose tolerance. *N Engl J Med* 2001;**344:**1343–50.

[59] Hu FB. Dietary pattern analysis: a new direction in nutritional epidemiology. *Curr Opin Lipidol* 2002;**13:** 3–9.

[60] Yu Z, Song G, Guo Z *et al*. Changes in blood pressure, body mass index, and salt consumption in a Chinese population. *Prev Med* 1999;**29:**165–72.

[61] Wang Y, Mi J, Shan XY, Wang QJ, Ge KY. Is China facing an obesity epidemic and the consequences? The trends in obesity and chronic disease in China. *Int J Obes (Lond)* 2006.