Public Health Nutrition: 24(15), 4918–4928 doi:10.1017/S1368980020004826

Association between dietary patterns and risk of breast cancer in Chinese female population: a latent class analysis

Shang Cao[1], Shurong Lu[2], Jinyi Zhou[2], Zheng Zhu[2], Wei Li[1], Jian Su[2], Hao Yu[2], Wencong Du[2], Lan Cui[2], Yunqiu Dong[2,3], Yun Qian[2,3], Ming Wu[1,2,†] and Pingmin Wei[1,*,†]

[1]Department of Epidemiology and Health Statistics, Southeast University, Dingjiaqiao Road 87, Nanjing 210009, Jiangsu, China: [2]Department of Chronic Disease Control, Jiangsu Provincial Center for Disease Control and Prevention, Nanjing, China: [3]Department of Health Promotion and Chronic Non-Communicable Disease Control, Wuxi Center for Disease Control and Prevention, Wuxi, China

Submitted 23 March 2020: Final revision received 8 November 2020: Accepted 24 November 2020: First published online 1 December 2020

Abstract

Objective: To determine if specific dietary patterns are associated with breast cancer (BC) risk in Chinese women.
Design: Latent class analysis (LCA) was performed to identify generic dietary patterns based on daily food-frequency data.
Setting: The Chinese Wuxi Exposure and Breast Cancer Study (2013–2014).
Participants: A population-based case–control study (695 cases, 804 controls).
Results: Four dietary patterns were identified, Prudent, Chinese traditional, Western and Picky; the proportion in the controls and cases was 0·30/0·32/0·16/0·23 and 0·29/0·26/0·11/0·33, respectively. Women in Picky class were characterised by higher extreme probabilities of non-consumption of specific foods, the highest probabilities of consumption of pickled foods and the lowest probabilities of consumption of cereals, soya foods and nuts. Compared with Prudent class, Picky class was associated with a higher risk (OR = 1·42, 95 % CI 1·06, 1·90), while the relevant association was only in post- (OR = 1·44, 95 % CI 1·01, 2·05) but not in premenopausal women. The Western class characterised by high-protein, high-fat and high-sugar foods, and the Chinese traditional class characterised by typical consumption of soya foods and white meat over red meat, both of them showed no difference in BC risk compared with Prudent class did.
Conclusions: LCA captures the heterogeneity of individuals embedded in the population and could be a useful approach in the study of dietary pattern and disease. Our results indicated that the Picky class might have a positive association with the risk of BC.

Keywords
Breast cancer
Dietary patterns
Latent class analysis
Epidemiology
Case–control

A multitude of in-vitro studies had reported the effect of individual-specific dietary components on breast cancer (BC)[1]. Nutritional studies have historically been focusing on specific nutrients or foods in isolation and oversimplified the complexity of foods[2,3]. The high degree of intercorrelation among various nutrients and foods makes it difficult to attribute effects to a single independent component; the interpretation and application of results were limited[4,5]. Now in nutrition epidemiology, the critical concept of food synergy has been convinced that nutrients exist in a purposeful biological sense in food, delivering their combinations reflect biological functionality.

Creating dietary patterns that inherently account for interactions among micronutrients and estimate overall dietary effects may provide a more robust approach for determining associations between disease and diet[6,7]. In general, dietary patterns are typically derived using two main approaches, 'a priori approach' by using a predefined dietary pattern and fitting the data into the indices, as Diet Quality Index[8,9], or 'a posteriori approach' by data-driven statistical reduction techniques explores dietary patterns, as cluster analysis, factor analysis (FA) and principal component analysis[9,10]. Both 'a priori' and 'a posteriori' approaches had been widely used for defining dietary patterns associated with various health outcomes. As Mediterranean dietary pattern measured by the compliance

†These authors contributed equally to this work.

*Corresponding author: Email mpw1963@126.com
© The Author(s), 2021. Published by Cambridge University Press on behalf of The Nutrition Society

https://doi.org/10.1017/S1368980020004826 Published online by Cambridge University Press

with 'a priori' defined dietary indices was associated with reducing BC risk[11,12], the Western dietary pattern derived by FA was linked with increasing the risk of BC[13–15]. Although various methods have been developed to explore dietary patterns in different populations, there were still some challenges in accurately identifying dietary patterns[16–18], including reducing the complex multidimensional nutritional data down to meaningful observed dietary patterns, dealing with the heterogeneity of individuals embedded in the studied population and classifying each participant into a specific dietary pattern[19,20].

Latent class analysis (LCA) is a person-centred, data-driven analytic approach[18,21]. LCA identifies the patterns of relations among a set of observed variables and classifies similar individuals into specific latent classes, which leads subjects to highly similar in class but uniquely different from the members of other classes[22]. Compared with 'variable-oriented' approaches, as FA/principal component analysis, that characterise the overall sample, 'person-centred' approaches model distinct configurations of heterogeneity within the sample, which could identify distinct, unknown patterns in subtypes of individuals from multidimensional data, capturing heterogeneity within and between groups. Moreover, LCA allows adjustment for covariates, quantifies member class probability and assessment of the goodness of fit[23,24].

Therefore, the study aims to use LCA to identify distinct classes of dietary patterns in Chinese women and evaluate whether specific dietary patterns are associated with BC risk and whether these associations will be affected by menopause status.

Methods

Study design and subjects

Chinese Wuxi Exposure and Breast Cancer Study (2013–2014) is a population-based case–control study. The subjects were all women who lived in Wuxi city, Jiangsu Province, China, for more than 5 years, as previously reported[25]. According to the cancer registration system, newly diagnosed BC patients within 1 year were selected as the case group. All cases were identified according to the International Classification of Diseases for Oncology (ICD-10, code C50). Patients with secondary or recurrent BC were excluded. For those with multiple incident cancers, we only included those with BC as the first diagnosed original malignancy. Controls were selected from the local population registry system and matched to the cases by the same residence area and age (range of ±5 years), excluding individuals with any cancer history.

From November 2013 to November 2014, a total of 1410 newly diagnosed BC cases were identified, 1072 cases meeting the inclusion criteria and 818 of them were recruited in the current study, the response rate was 76·3 % (818/1072). Moreover, 1072 controls were selected and 935 of them participated, with a response rate of 87·2 % (935/1072), conducted as a frequency-matched case–control study (Fig. 1). The study protocol was approved by the Institutional Review Boards of Jiangsu CDC, and written informed consent was obtained from all subjects.

Data collection

Demographic, lifestyle characteristics, menstrual and reproductive events, dietary intake, disease history and physical activity-related data were collected during the person-to-person interviews conducted by trained interviewers. Anthropometric measures were obtained by trained personnel following a standard protocol. The usual diet was assessed by a validated, semi-quantitative FFQ, which included 149 items along with the recipes commonly used in China, a detailed description given[26]. Physical activity was measured by referencing the Global Physical Activity Questionnaire[27]. Nutrient and energy intake were calculated through the Chinese Food Composition Database (2018, 6th version).

Dietary patterns derived by latent class analysis

Dietary intake assessment included whether the food was consumed, consumption frequency (times per day/week/month/year) and the average amount of food consumption at each time. The 149 food items in the FFQ were classified into eighteen predefined food groups based on similarities in nutrient profile and culinary usage, including rice/flour, cereals, fried foods, red meat, poultry, aquatic products, eggs, milk, fruits, vegetables, soya foods, nuts, cakes, sugar strengthened beverage, fresh juice, soft drink, pickled foods and coffee.

LCA identifies the number of 'latent' classes that describe the association between manifest variables. LCA estimates two key sets of parameters for categorical outcomes and covariates, respectively[28]: conditional probabilities (or called response probability) of observer indicators under a given class (e.g., the probability of red meat consumption among women adhering to Western class) and the regression coefficients predicting class membership (e.g., to differentiate between two subjects under the same energy intake but with different types of food consumption per day, as 250 g/d fire food and 1000 g/d rice respectively, with same 8368 kJ/d (2000 kcal/d)). Two sets of parameters were given to predict the posterior probability of belonging to each class for each subject[29].

Since the probability of food items consumption often had a typical spike at 0 for non-consumers and constrained between 0 and 1 for consumers, we categorised the derived food group consumption into four levels: tertiles of non-zero consumption and no consumption (calculated from controls). Because there were < 20 % of women consumed sugar strengthened beverage, fresh juice, soft drink or coffee, we set the consumption of these foods as binary variables (consumed or no). While rice/flour was consumed almost ubiquitously, there were only tertiles of consumption and no non-consumption category.

https://doi.org/10.1017/S1368980020004826 Published online by Cambridge University Press

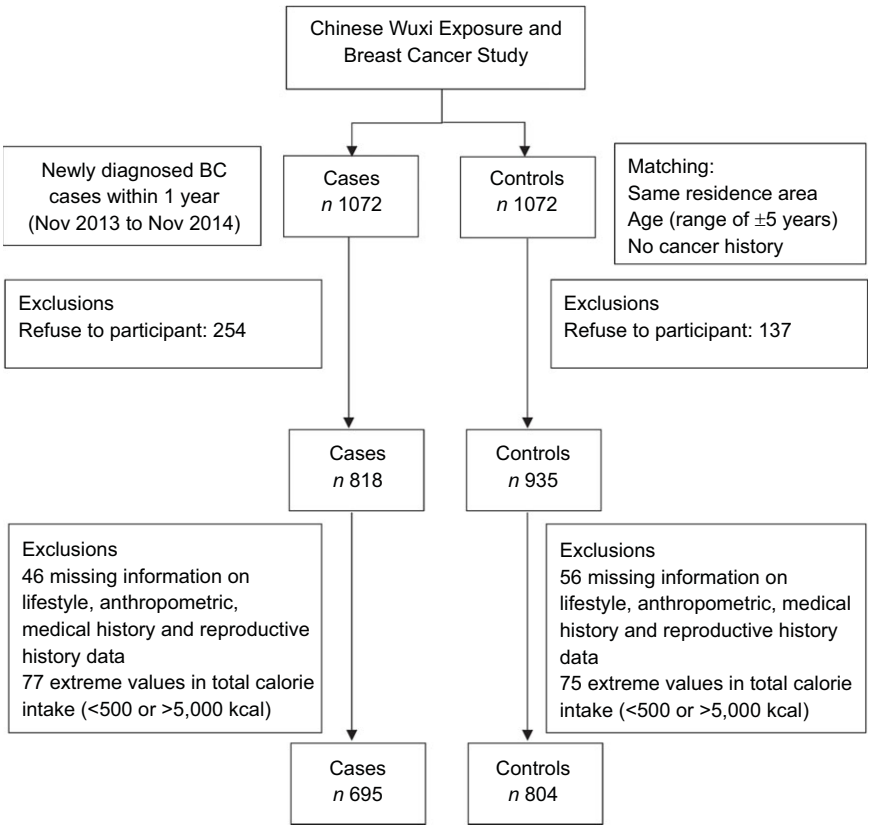

Fig. 1 Flow chart of the participants, the Chinese Wuxi Exposure and Breast Cancer Study

LCA was used to derive dietary patterns of food groups based on the dietary data from controls, adjusted for energy intake (kcal/d). The dietary classes were interpreted and named according to the conditional probabilities of food group intake. The number of classes was determined using the Bayesian information criterion, Lo-Mendel-Rubin likelihood ratio (LMR) test and entropy value to identify the best-fitting model with statistical fit balance and parsimony. Finally, we predicted the probabilities of class membership for controls and cases, assigning them to the specific latent class (dietary pattern) for which the probability is the highest.

Statistical analysis

All analyses were stratified by menopausal status at diagnosis for cases or enrollment for controls. Women were considered to be postmenopausal as an absence of menstruation in the past 12 months.

Associations between exposures (dietary pattern as a nominal predictor) and the outcomes (BC) were estimated in terms of adjusted OR and corresponding 95 % CI using the logistic regression model. All models were adjusted for age at diagnosis for cases or enrollment for controls (by years), area (urban, rural), education (ordered as illiterate and primary, middle and high school, university and above), tobacco smoking (no or yes: including smoking and second-hand smoking ≥ 3d/week), tea intake (no or yes: ≥ 3 d/week), alcohol intake (no or yes: ≥ 3 d/week), moderate physical activity (min/d), oral contraceptives use (no or yes: current use or ever use), hormone replacement therapy (HRT) (no or yes: current use or ever use), family history of BC (no or yes: in a first-degree relative), history of benign breast disease (no or yes: including lactation mastitis, plasma cell mastitis, cyclomastopathy, fibroadenoma of breast, galactocele), age at menarche (by years), parity (ordered as 0, 1, 2 or ≥ 3), age at first full-term delivery (by years), breast-feeding (no or yes), height (by cm), BMI (kg/m^2) and energy intake (kcal/d). Postmenopausal stratification analysis was further adjusted for the menopausal age (by years). Furthermore, to examine whether the association between dietary patterns and BC risk was affected by well-established or suspected non-dietary BC risk factors, we conducted stratified analysis and interaction test by selected covariates which can cause a change in the OR of interest by at least 10 %, including age, education, tobacco smoking, alcohol intake, HRT, age at menarche, age at first full-term delivery, menopausal age, family history of BC, history of benign breast disease, parity, energy intake, BMI and height (stratification based on the median of controls distribution). The $P_{\text{for interaction}}$ was calculated by the likelihood ratio test.

Finally, to assess the impact of classification quality, we also excluded women with low predicted probabilities of the class membership (< (K–1)/K in LCA with K classes) for their assigned dietary class as a sensitivity analysis.

https://doi.org/10.1017/S1368980020004826 Published online by Cambridge University Press

The LCA was conducted using MPLUS (V8.3; Muthén & Muthén), and other statistical analyses were conducted using R version 4.0.0 (The R Project for Statistical Computing, USA; https://www.r-project.org/). All P values quoted were 2-sided, and $< 0{\cdot}05$ was considered as statistically significant.

Results

Of the participants interviewed (818 cases, 935 controls), we excluded seventy-seven cases and seventy-five controls because of extreme values in total energy intake (< 500 or > 5000 kcal) and forty-six cases and fifty-six controls missing information on adjusting covariant variables. No significant difference among demographic characteristics was found between excluded participants and remained ones. The results presented here were based on 695 cases and 804 controls who have complete information on FFQ and possible covariates included for an adjustment. The education level of the cases lower than that of the controls, the overweight rate, family history of BC and history of benign breast disease of cases was higher than that of the controls ($P < 0{\cdot}01$) (Table 1).

Dietary patterns derived by latent class analysis

Latent class models were fitted for two to six classes, and finally, four dietary pattern classes were chosen. Because when the LCA model retains four classes, the value of Bayesian information criterion is the smallest, the Lo-Mendel-Rubin likelihood ratio test reaches a significant level and the entropy value (0·836) is ideal. The proportion of each class after dividing into four categories is also balanced. Therefore, based on the balance of statistical fit and parsimony, we believe that the four-class model is appropriate (see online supplementary material, Supplemental Table 1). Figure 2 shows the conditional probabilities of Chinese women taking each food group in each class. A food group in high (third tertile) consumption (Fig. 2a) had a probability close to 1 in a given latent class, suggesting women in that class were likely to take more of that specific food. The food group in no consumption (Fig. 2b) with a probability close to 1 indicated that women took food less often (see online supplementary material, Supplemental Table 2). We named the four chosen classes as Prudent diet consumers, Western diet consumers, Chinese traditional diet consumers and Picky diet consumers.

Characteristics of dietary patterns

The Prudent class was characterised by a high probability of consuming healthy foods such as cereals, aquatic products, fruits, vegetables, soya foods and nuts. The Chinese traditional class was featured by the preference of white meat (as poultry) over red meat and the general willingness to take soya foods, with the lowest probabilities in non-consumption of the specific foods. Western class shows the highest probability of consuming a high-protein, high-fat and high-sugar foods such as fried meat or eggs, cakes, soft drinks, coffee, as well as soya foods. Compared with other classes, women in the Picky class were characterised by higher extreme probabilities of non-consumption of specific foods. In addition, women in the Picky class showed the highest probabilities of consumption of pickled foods and the lowest probabilities of consumption of cereals, soya foods and nuts (see online supplementary material, Supplemental Table 3). Overall, by conditional probability, 29·0, 31·9, 15·7 and 23·4 % of women were characterised by Prudent class, Chinese traditional class, Western class and Picky class, respectively. Women characterised by Western class demonstrated a significantly higher amount of energy intake than those the other three classes did (Western: $2226{\cdot}71 \pm 414{\cdot}84$ kcal/d, Chinese traditional: $1630{\cdot}53 \pm 283{\cdot}57$ kcal/d, Prudent: $1860{\cdot}84 \pm 280{\cdot}99$ kcal/d, Picky: $1555{\cdot}98 \pm 272{\cdot}89$ kcal/d). Additionally, we compared the socio-demographic characteristics within the identified latent classes and found significant differences between the classes (see online supplementary material, Supplemental Table 4).

Association between dietary pattern and breast cancer risk

Regarding the Prudent class as the reference group, the Picky class showed an independent risk effect on BC with an OR of 1·46 (95 % CI 1·01, 2·05) among postmenopausal women, whereas Western class (OR = 0·87, 95 % CI 0·54, 1·43) and Chinese traditional class (OR = 0·90, 95 % CI 0·63, 1·29) showed no difference, while no relevant association was found in premenopausal women (Table 2). Because the associations of BC risk with LCA-driven dietary patterns were not found among premenopausal women, stratified analysis and interaction tests were restricted only to the postmenopausal women. As shown in Table 3, we found that the associations of dietary patterns with BC risk were affected by some well-established or suspected non-dietary BC risk factors, that is, the Picky class was strongly associated with increased risk of BC among women who were age < 60 (years), drink alcohol, never use of HRT, age of menarche ≥ 16 (years), age at first full-term delivery ≥ 25(years), menopausal age ≥ 55, had a history of benign breast disease, parity < 2 or height < 155 cm.

Sensitivity analysis

The sensitivity analysis of dietary classification quality was assessed by excluding women (20·98 % of controls, 27·48 % of cases) with a low predicted probability (< 0·75) of the class membership. However, none of the associations between dietary classes and BC risk changed substantially.

Finally, we analysed the impact of selection bias caused by non-responders on the results; the possible selection bias would not change the existing conclusions. Through telephone follow-up, we investigated

https://doi.org/10.1017/S1368980020004826 Published online by Cambridge University Press

https://doi.org/10.1017/S1368980020004826 Published online by Cambridge University Press

Table 1 Characteristics of cases and controls for analysis, stratified by menopausal status

	All					Premenopausal					Postmenopausal				
	Case (*n* 695).		Control (*n* 804)			Case (*n* 224)		Control (*n* 332)			Case (*n* 471)		Control (*n* 472)		
Variable	*n*	%	*n*	%	*P**	*n*	%	*n*	%	*P**	*n*	%	*n*	%	*P**
Age at enrollment (years)															
< 50	282	40·58	339	42·16	0·22	201	89·7	288	86·7	0·31	81	17·2	51	10·8	0·01
50–60	180	25·90	228	28·36		18	8·0	39	11·7		162	34·4	189	40·0	
> 60	233	33·53	237	29·48		5	2·2	5	1·5		228	48·4	232	49·2	
Mean	55·07		54·42		0·27	44·65		44·79		0·81	60·03		61·20		0·05
SD	11·27		11·28			6·69		6·96			9·47		8·45		
Area															
Urban	380	54·68	454	56·47	0·49	119	53·13	182	54·82	0·69	261	55·41	272	57·63	0·49
Rural	315	45·32	350	43·53		105	46·88	150	45·18		210	44·59	200	42·37	
Educational level															
Illiterate and primary	223	32·09	199	24·75	< 0·01	35	15·6	24	7·2	< 0·01	188	39·9	175	37·1	0·13
Middle and high school	301	43·31	290	36·07		124	55·4	126	38·0		177	37·6	164	34·7	
University and above	171	24·60	315	39·18		65	29·0	182	54·8		106	22·5	133	28·2	
Tobacco smoking															
No	291	41·87	348	43·29	0·58	76	33·9	126	38·0	0·33	215	45·6	222	47·0	0·67
Yes	404	58·13	456	56·72		148	66·1	206	62·0		256	54·4	250	53·0	
Alcohol intake															
No	645	92·81	738	91·79	0·46	205	91·52	293	88·25	0·22	440	93·42	445	94·28	0·58
Yes	50	7·19	66	8·21		19	8·48	39	11·75		31	6·58	27	5·72	
Tea intake															
No	563	81·01	619	76·99	0·06	179	79·91	242	72·89	0·06	384	81·53	377	79·87	0·52
Yes	132	18·99	185	23·01		45	20·09	90	27·11		87	18·47	95	20·13	
Moderate physical activity (min/d)															
Mean	16·42		16·33		0·90	15·19		13·89		0·22	17·00		18·05		0·28
SD	14·38		13·60			13·00		11·54			14·98		14·65		
Oral contraceptives use															
No	132	18·99	147	18·28	0·73	41	18·3	63	19·0	0·84	91	19·3	84	17·8	0·55
Yes	563	81·00	657	81·72		183	81·7	269	81·0		380	80·7	388	82·2	
Hormone replacement therapy															
No	659	94·82	774	96·27	0·17	212	94·6	316	95·2	0·78	447	94·9	458	97·0	0·10
Yes	36	5·18	30	3·73		12	5·4	16	4·8		24	5·1	14	3·0	
Family history of breast cancer															
No	621	89·35	764	95·02	< 0·01	197	87·9	313	94·3	0·01	424	90·0	451	95·6	< 0·01
Yes	74	10·65	40	4·98		27	12·1	19	5·7		47	10·0	21	4·4	
History of begin disease															
No	424	61·01	550	68·41	< 0·01	113	50·4	183	55·1	0·28	311	66·0	367	77·8	< 0·01
Yes	271	38·99	254	31·59		111	49·6	149	44·9		160	34·0	105	22·2	
Age at menarche															
10–14	203	29·21	244	30·35	0·89	94	42·0	148	44·6	0·67	109	23·1	96	20·3	0·58
15–16	267	38·42	305	37·94		90	40·2	121	36·4		177	37·6	184	39·0	
17–22	225	32·37	255	31·72		40	17·9	63	19·0		185	39·3	192	40·7	
Mean	15·65		15·64		0·97	14·93		14·95		0·91	15·99		16·13		0·23
SD	1·83		1·88			1·68		1·65			1·81		1·89		

https://doi.org/10.1017/S1368980020004826 Published online by Cambridge University Press

Table 1 *Continued*

	All					Premenopausal					Postmenopausal				
	Case (*n* 695).		Control (*n* 804)			Case (*n* 224)		Control (*n* 332)			Case (*n* 471)		Control (*n* 472)		
Variable	*n*	%	*n*	%	*P**	*n*	%	*n*	%	*P**	*n*	%	*n*	%	*P**
Parity															
0	9	1·29	11	1·37	0·61	2	0·9	5	1·5	0·30	7	1·5	6	1·3	0·74
1	383	55·11	467	58·08		171	76·3	269	81·0		212	45·0	198	41·9	
2	219	31·51	228	28·36		47	21·0	50	15·1		172	36·5	178	37·7	
≥ 3	84	12·09	98	12·19		4	1·8	8	2·4		80	17·0	90	19·1	
Age at first full-term delivery															
≤ 25	349	50·22	406	50·50	0·18	129	57·6	183	55·1	0·05	220	46·7	223	47·2	0·81
25–29	317	45·61	378	47·01		88	37·4	147	44·3		229	49·8	231	48·9	
> 29	29	4·17	20	2·49		7	3·1	2	0·6		22	4·7	18	3·8	
Mean	24·59		24·56		0·84	24·11		24·38		0·16	24·82		24·69		0·47
SD	2·65		2·49			2·40		2·10			2·74		2·73		
Breast-feeding															
No	237	34·10	248	30·85	0·18	48	21·4	58	17·5	0·24	189	40·1	190	40·3	0·97
Yes	458	65·90	556	69·15		176	78·6	274	82·5		282	59·9	282	59·7	
Height (cm)															
Mean	156·90		156·76		0·62	158·05		158·75		0·14	156·35		155·35		< 0·01
SD	5·29		5·64			5·26		5·50			5·22		5·32		
BMI															
Underweight (< 18·5)	15	2·16	13	1·62	< 0·01	7	3·1	7	2·1	0·02	8	1·7	6	1·3	< 0·01
Normal (18·5–23·9)	266	38·27	406	50·50		104	46·4	193	58·1		175	37·2	221	46·8	
Overweight (> 24)	414	59·57	385	47·89		113	50·4	132	39·8		288	61·1	245	51·9	
Mean	24·95		24·08		< 0·01	24·44		23·63		0·01	25·20		24·39		< 0·01
SD	3·54		3·14			3·62		3·12			3·87		3·12		
Energy intake (kcal/d)															
Mean	1733·61		1782·24		0·14	1765·11		1873·51		0·06	1718·63		1718·05		0·99
SD	642·77		639·87			645·95		676·68			641·40		605·19		
Menopause age															
Mean											49·23		49·44		0·49
SD											4·30		5·00		

*P-values are calculated based on the χ^2 test or *t* test.

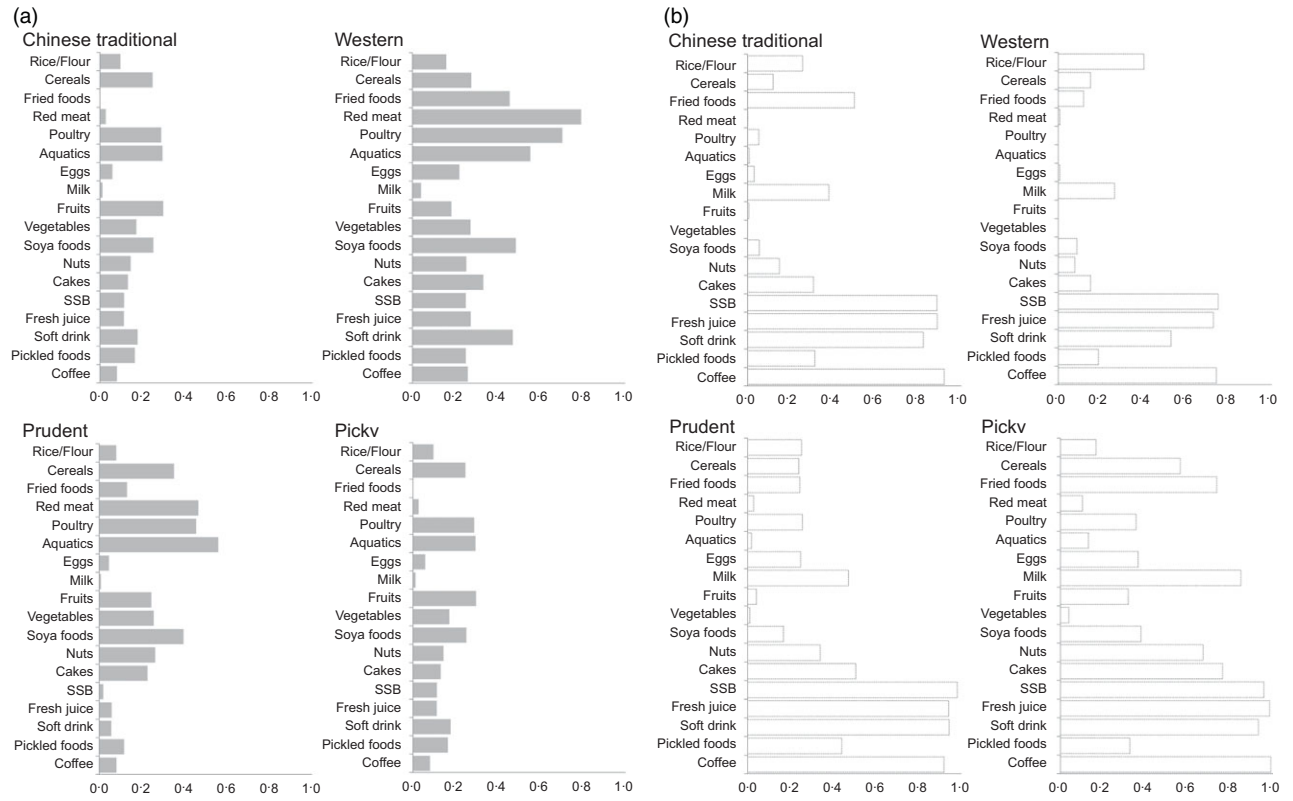

Fig. 2 Food consumption level conditional probabilities of dietary pattern class. (a) High consumption conditional probability of dietary pattern class. (b) No conditional consumption probability of dietary pattern class

Table 2 OR and 95 % CI for the association of latent class analysis (LCA)-derived dietary patterns and breast cancer risk

Dietary patterns	Case/Control	Crude OR*	95 % CI	Adjusted OR*,†	95 % CI	$P_{interaction}$ Diet × menopause‡
All (*n* 1499)						
Prudent	204/241	1·00	Reference	1·00	Reference	0·0131
Chinese traditional	184/257	0·85	0·65, 1·10	0·86	0·65, 1·14	
Western	79/125	0·75	0·53, 1·05	0·76	0·53, 1·09	
Picky	228/181	1·49	1·14, 1·95	1·42	1·06, 1·90	
Stratified by menopausal status						
Premenopausal (*n* 556)						
Prudent	66/91	1·00	Reference	1·00	Reference	
Chinese traditional	65/115	0·78	0·51, 1·21	0·76	0·47, 1·23	
Western	29/74	0·54	0·32, 0·92	0·62	0·35, 1·11	
Picky	64/52	1·70	1·05, 2·75	1·45	0·84, 2·48	
Postmenopausal (*n* 943)						
Prudent	138/150	1·00	Reference	1·00	Reference	
Chinese traditional	119/142	0·91	0·65, 1·27	0·90	0·63, 1·29	
Western	50/51	1·07	0·68, 1·68	0·87	0·54, 1·43	
Picky	164/129	1·38	1·00, 1·92	1·44	1·01, 2·05	

*OR were derived from the logistic regression model.
†Adjusted for age, area, education, tobacco smoking, tea intake, alcohol intake, moderate physical activity, oral contraceptives use, hormone replacement therapy, family history of breast cancer, history of benign breast disease, age at menarche, parity, age at first full-term delivery, breast-feeding, height, BMI, energy intake, postmenopausal additional adjusted for the menopausal age.
‡$P_{for\ interaction}$ between dietary patterns and menopause was derived from a likelihood ratio test.

the reasons for non-respondents and fill them with random sampling from existing samples based on their characteristics, using fully conditional specification multivariate imputation by the chained equations method[30] (see online supplementary material, Supplemental 'Selection bias analysis').

Discussion

In the current study, we applied a novel approach LCA to identify the generic dietary patterns in the Chinese female population and evaluated their associations with the risk of BC. We found that the Picky class contributed an additional

https://doi.org/10.1017/S1368980020004826 Published online by Cambridge University Press

https://doi.org/10.1017/S1368980020004826 Published online by Cambridge University Press

Table 3 Associations between postmenopausal breast cancer risk and latent class analysis (LCA)-derived dietary patterns, stratified by suspected non-dietary breast cancer risk factors

		Dietary pattern, adjusted*							
			Chinese traditional		Western		Picky		
Subgroup	Cases/Controls	Prudent	OR	95 % CI	OR	95 % CI	OR	95 % CI	$P_{for\ interaction}$†
Age (years)									0·0980
< 60	225/213	1·0	0·91	0·53, 1·55	0·93	0·47, 1·83	2·24	1·26, 4·01	
≥ 60	246/259	1·0	1·05	0·62, 1·77	0·74	0·33, 1·63	1·18	0·72, 1·94	
Education (years)									0·4774
< 7	188/175	1·0	1·12	0·63, 1·99	1·11	0·49, 2·50	1·21	0·72, 2·03	
> 7	283/297	1·0	0·82	0·54, 1·24	1·05	0·61, 1·81	1·55	1·00, 2·39	
Tobacco smoking									0·7529
No	215/222	1·0	0·77	0·47, 1·26	1·10	0·51, 2·35	1·21	0·76, 1·94	
Yes	256/250	1·0	1·06	0·67, 1·67	1·07	0·61, 1·90	1·57	0·99, 2·47	
Alcohol									0·6378
No	440/445	1·0	0·89	0·63, 1·26	1·03	0·64, 1·65	1·43	1·02, 2·00	
Yes	31/27	1·0	1·02	0·23, 4·47	1·25	0·22, 7·08	0·74	0·16, 3·39	
Hormone replacement therapy									0·4616
No	447/458	1·0	0·97	0·69, 1·37	1·08	0·67, 1·73	1·44	1·03, 2·01	
Yes	24/14	1·0	0·25	0·04, 1·47	0·69	0·12, 4·08	1·67	0·15, 18·88	
Age of menarche (years)									0·6776
< 16	177/173	1·0	0·81	0·46, 1·42	0·80	0·37, 1·76	1·04	0·55, 1·98	
≥ 16	294/299	1·0	0·99	0·62, 1·59	0·90	0·47, 1·72	1·60	1·03, 2·47	
Age at first full-term delivery (years)									0·1053
< 25	297/304	1·0	0·75	0·47, 1·20	0·74	0·38, 1·43	1·15	0·72, 1·83	
≥ 25	174/168	1·0	1·08	0·56, 2·08	1·08	0·47, 2·46	2·54	1·34, 4·83	
Menopausal age (years)									0·0724
< 55	206/193	1·0	1·11	0·62, 2·00	1·13	0·53, 2·44	1·25	0·70, 2·23	
≥ 55	265/279	1·0	0·77	0·48, 1·24	0·64	0·32, 1·25	1·77	1·10, 2·85	
Family history of breast cancer									0·8357
No	311/367	1·0	0·91	0·64, 1·29	1·08	0·67, 1·75	1·39	0·99, 1·95	
Yes	160/105	1·0	0·71	0·18, 2·84	0·54	0·12, 2·52	1·35	0·30, 6·13	
History of benign breast disease									0·0085
No	424/451	1·0	1·26	0·81, 1·94	1·08	0·58, 2·01	1·45	0·95, 2·23	
Yes	47/21	1·0	0·47	0·23, 0·97	0·38	0·15, 0·98	2·45	1·06, 5·62	
Parity									0·2537
< 2	219/204	1·0	1·24	0·72, 2·15	1·38	0·68, 2·80	1·88	1·03, 3·44	
≥ 2	252/268	1·0	0·77	0·46, 1·30	0·48	0·22, 1·05	1·27	0·79, 2·05	
Energy (kcal)									0·1133
< 1655	256/236	1·0	0·75	0·44, 1·27	0·63	0·26, 1·52	1·56	0·94, 2·58	
≥ 1655	215/236	1·0	1·06	0·62, 1·82	0·92	0·49, 1·71	0·99	0·56, 1·75	
BMI (kg/m^2)									0·7609
< 24	183/227	1·0	0·76	0·43, 1·34	0·79	0·38, 1·65	1·48	0·86, 2·54	
≥ 24	288/245	1·0	1·07	0·65, 1·76	0·94	0·46, 1·89	1·39	0·84, 2·31	
Height (cm)									0·6745
< 155	169/222	1·0	1·05	0·58, 1·90	0·86	0·35, 2·14	2·01	1·14, 3·56	
≥ 155	302/250	1·0	0·86	0·53, 1·38	0·81	0·44, 1·50	1·27	0·78, 2·06	

*Adjusted for age, area, education, tobacco smoking, tea intake, alcohol intake, moderate physical activity, oral contraceptives use, hormone replacement therapy, family history of breast cancer, history of benign breast disease, age at menarche, parity, age at first full-term delivery, breast-feeding, height, BMI, energy intake, menopausal age.
†$P_{for\ interaction}$ between dietary patterns and non-dietary BC risk factors was derived from a likelihood ratio test.
Note: in this study, not all the food groups had a third tertile level (e.g., sugar strengthened beverage, fresh juice, etc. were binary) as the title states. And for food groups that had four levels, there is one non-consumption category.

risk to BC (OR = 1·42, 95 % CI 1·06, 1·90), while the relevant association was only in post- (OR = 1·44, 95 % CI 1·01, 2·05) but not in premenopausal women. The Western class and Chinese traditional classes were not associated with beneficial or adverse effects on BC risk compared with the Prudent class.

In nutritional epidemiology, data-driven methods such as FA/principal component analysis were widely used for nutritional data reduction, but the challenges in accurately identifying dietary patterns across population still exist. For these 'variable-oriented' methods, food items are grouped according to the degree of association between each other, operating by partitioning variance between measured variables[31,32]. However, in most of the cases, the asymmetric distribution of variables reflecting food consumption caused by the heterogeneity of individuals in the diet will hinder the full capture of generic dietary patterns in the studied population[16,17]. Because this type of method performs on the square of simple correlation coefficients between variables, skewed distribution results in the sum of squares of simple correlation coefficients between variables being much smaller than the sum of squares of partial correlation coefficients, and the lower variance contribution will make it challenging to capture the information about the relationship between the variables of interest[33].

Therefore, considering that the source of heterogeneity is from the individuals of the studied population rather than the diet measurement variables themselves, studying 'person-centred' instead of 'variable-oriented' may be more effective[34]. We compared the results of LCA and FA based on the same data sets (online supplementary material, Supplemental 'Comparison between LCA and FA'). We found that the classification of dietary patterns was roughly similar to the previous study based on the FA approach, which demonstrates LCA and FA identified similar dietary patterns when presented with the same data set. However, heterogeneity embedded in the study population leads to an unreliable result of FA; a low original variance (45·21 %) may affect obtained findings. In contrast, LCA with an ideal entropy value (0·836) ensures the accuracy of classification (> 90 %). Additionally, we compared the alternate Mediterranean Diet score to the highly data-driven LCA results; as 'a priori' approach, it could better capture specific dietary characteristics under an identified actual dietary pattern. We found that the alternate Mediterranean Diet score indices seemed similar to the Prudent dietary pattern in terms of its correlation with specific foods (online supplementary material, Supplementary 'Comparison between LCA and DQI'). The comparison demonstrates that under the premise that LCA is conducive to identifying the heterogeneity in different subpopulations, it also has a good performance in understanding the combination of food consumption and capturing the diet characteristics in a specific dietary pattern.

The results of the association between dietary patterns and BC risk were not consistent in Western population[1,13–15]. Inverse associations with Prudent dietary pattern and positive associations with the Western dietary pattern of BC risk have been found in many studies. However, some studies also reported contradictory findings[1,13–15]. Similarly, the current research on Asian women's dietary patterns and BC risk has not reached a consistent conclusion[35–39]. In our study, no difference between BC risk and Prudent class, Western class or Chinese traditional class was found. Although the characteristics of the three dietary patterns were different, we have found some commonalities. Compared with Picky class, the probability of soya foods intake in Prudent, Chinese and Western was relatively higher. Soya isoflavones may reduce the risk of BC by preferentially binding the oestrogen-dependent mechanism of oestrogen receptor-b relative to oestrogen receptor-α[40], as well as the oestrogen-independent mechanism of inhibiting the nuclear transcription factor κB DNA binding activity and the Akt signalling pathway[41]. We did not find Western class had a positive association with BC risk in the current study, a possible reason for that might be the relatively high consumption of polyphenols foods such as soya foods and nuts in the Chinese women of Western class. Besides, Chinese women's average red meat intake (19·32 g/1000 kcal/d) was much lower than that among Asian Americans (34·5 g/1000 kcal/d)[42], which was different from a typical Western diet pattern among Western populations.

What deserves attention in the current study is the Picky class. We first found women's compliance with this dietary pattern was with a higher BC risk among postmenopausal women compared with those of Prudent class. Women in Picky class were characterised by higher extreme probabilities of non-consumption of specific foods, the highest probabilities in consumption of pickled foods and the lowest probabilities in consumption of cereals, soya foods and nuts. Therefore, we suspected that the high BC risk of Picky class might come from an imbalance diet that could lead to loss of certain vital nutrients and high consumption of pickled foods that are prone to inflammation.

For BC, the observed heterogeneity of risk affected by menopausal status was particularly substantial, as it hinted at relative contributions of oestrogens, progesterone, insulin and insulin-like growth factor 1 (IFG-1) in mediating the association[43]. The meaningful findings were concentrated in the postmenopausal women population; while during the postmenopausal period, oestrogens appear to be a dominant driver. Therefore, a potential biological mechanism may explain the null finding among premenopausal women: the ovaries are a predominant site of oestrogen synthesis in the premenopausal period; the additional contribution of diet factors to the circulating pool of oestrogens (i.e., oestrone, oestradiol, oestriol) may be negligible. Not only the amount of oestrogens from adipocytes is far smaller, but also the form of oestrogens (i.e., oestrone rather than

https://doi.org/10.1017/S1368980020004826 Published online by Cambridge University Press

oestradiol) is less biologically potent[44]. However, the conclusion of the association between dietary patterns and breast cancer differed by the menopausal status was not consistent[45]; also, this meaningless premenopausal association may be because of the sample size, a small sample size reduces the power of the study and increase the margin of error, more in-depth and cross-validation studies are needed. Furthermore, among postmenopausal women, we found that the association between diet and BC risk could be affected by some well-established or potential non-dietary risk factors, the interaction between these factors and compliance with Picky dietary pattern associated with BC risk appears to be complicated and some factors increased the risk of association between Picky dietary pattern and BC risk, but some others weakened this association. Notably, we found that the Picky pattern was at higher risk in non-alcohol drinkers and non-HRT users, implying that some strong independent risk factors for BC may mask the association between diet and BC risk. For example, oestrogens appear to be a dominant driver of BC in postmenopausal women. In the absence of excess oestrogens from the ovaries and HRT, variation in oestrogen levels because of the different dietary patterns may be sufficient to distinguish the risk of BC. In contrast to HRT users, exogenous oestrogens from HRT may raise plasma oestrogens to the extent that endogenous oestrogens from the diet have a little incremental effect. Further work should assess associations of BC risk and the concentrations of these nutrients in plasma, which may be more predictive for vivo situation and interpretative for disease risk and biological mechanism.

However, several limitations also should be noted. First, data were collected from a case–control study, which might be partially influenced by the biases inherent in case–control designs, included selection bias, recall bias, residual confounding and reverse causality. We only included newly diagnosed BC patients and design the question carefully to reduce the recall bias, and the dietary preference was collected based on composite measures in which it was less likely to cause information bias on specific foods/food groups. We try to reduce the influence of residual confounding on conclusions through stratified analysis. Also, we analysed the potential effect of selection bias in the current study, and it does not change the conclusion. Another limitation is the lack of information on the receptor status of breast tumours, which might lead to an underestimated impact of diet on BC risk. However, some studies reported that the associations between dietary patterns and BC risk did not change substantially by receptor status[46,47]. Lastly, the results obtained through LCA tend to be highly data-driven and require cross-validation with other independent samples in the future.

Conclusions

In conclusion, we found that LCA is a useful approach to capture dietary patterns within complex dietary data of high inter-individual variation and to derive interpretable dietary patterns suitable for associating with health outcomes. Our findings further support the hypothesis that the combinations of specific dietary factors protect against BC and may be involved in the mechanisms of action on breast carcinogenesis.

Acknowledgements

Acknowledgements: We appreciate all study participants for their contributions. The authors thank the entire data collection team. Incident BC cases and controls for the current study were collected by Wuxi Center for Disease Control and Prevention, Jiangsu Center for Disease Control and Prevention. *Financial support:* The current study was supported by the World Cancer Research Fund (2011/RFA/473) and Wuxi Young Medical Talents (QNRC035). *Conflict of interest:* No. *Authorship:* The authors' responsibilities were as follows: M.W., S.C., S.R.L. and P.M.W.: designed and conducted the study; J.Y.Z., Z.Z., L.W., J.S., H.Y., W.D., L.C., Y.Q.D. and Y.Q.: developed plant-based diet indices and data collection; S.C. and M.W.: performed the statistical analyses and drafted the manuscript; S.C., M.W., P.M.W. and Z.Z.: interpreted the data, critically revised the manuscript and had full responsibility for the analyses and interpretation of the data; S.C.: full access to all study data; and all authors: read and approved the final manuscript. *Ethics of human subject participation:* The current study was conducted according to the guidelines laid down in the Declaration of Helsinki and all procedures involving human subjects/patients were approved by the Jiangsu Center for Disease Control and Prevention ethical committee. Written informed consent was obtained from all subjects/patients. All participants signed written informed consent.

Supplementary material

For supplementary material accompanying this paper visit https://doi.org/10.1017/S1368980020004826

References

1. Gandini S, Merzenich H, Robertson C *et al.* (2000) Meta-analysis of studies on breast cancer risk and diet: the role of fruit and vegetable consumption and the intake of associated micronutrients. *Eur J Cancer* **36**, 636–646.
2. Michels KB, Mohllajee AP, Roset-Bahmanyar E *et al.* (2007) Diet and breast cancer: a review of the prospective observational studies. *Cancer* **109**, 2712–2749.
3. Kerr J, Anderson C & Lippman SM (2017) Physical activity, sedentary behaviour, diet, and cancer: an update and emerging new evidence. *Lancet Oncol* **18**, e457–e471.
4. Hu FB (2002) Dietary pattern analysis: a new direction in nutritional epidemiology. *Curr Opin Lipidol* **13**, 3–9.

https://doi.org/10.1017/S1368980020004826 Published online by Cambridge University Press

5. Tapsell LC (2014) Foods and food components in the Mediterranean diet: supporting overall effects. *BMC Med* **12**, 100.
6. Gleason PM, Boushey CJ, Harris JE *et al.* (2015) Publishing nutrition research: a review of multivariate techniques – part 3: data reduction methods. *J Acad Nutr Diet* **115**, 1072–1082.
7. Castello A, Buijsse B, Martin M *et al.* (2016) Evaluating the applicability of data-driven dietary patterns to independent samples with a focus on measurement tools for pattern similarity. *J Acad Nutr Diet* **116**, 1914–1924 e1916.
8. Kant AK (1996) Indexes of overall diet quality: a review. *J Am Diet Assoc* **96**, 785–791.
9. Ocke MC (2013) Evaluation of methodologies for assessing the overall diet: dietary quality scores and dietary pattern analysis. *Proc Nutr Soc* **72**, 191–199.
10. Newby PK & Tucker KL (2004) Empirically derived eating patterns using factor or cluster analysis: a review. *Nutr Rev* **62**, 177–203.
11. Sofi F, Cesari F, Abbate R *et al.* (2008) Adherence to Mediterranean diet and health status: meta-analysis. *BMJ* **337**, a1344.
12. Bloomfield HE, Koeller E, Greer N *et al.* (2016) Effects on health outcomes of a Mediterranean diet with no restriction on fat intake: a systematic review and meta-analysis. *Ann Intern Med* **165**, 491–500.
13. Albuquerque RC, Baltar VT & Marchioni DM (2014) Breast cancer and dietary patterns: a systematic review. *Nutr Rev* **72**, 1–17.
14. Brennan SF, Cantwell MM, Cardwell CR *et al.* (2010) Dietary patterns and breast cancer risk: a systematic review and meta-analysis. *Am J Clin Nutr* **91**, 1294–1302.
15. Grosso G, Bella F, Godos J *et al.* (2017) Possible role of diet in cancer: systematic review and multiple meta-analyses of dietary patterns, lifestyle factors, and cancer risk. *Nutr Rev* **75**, 405–419.
16. Edefonti V, Randi G, La Vecchia C *et al.* (2009) Dietary patterns and breast cancer: a review with focus on methodological issues. *Nutr Rev* **67**, 297–314.
17. Jacques PF & Tucker KL (2001) Are dietary patterns useful for understanding the role of diet in chronic disease? *Am J Clin Nutr* **73**, 1–2.
18. Uzhova I, Woolhead C, Timon CM *et al.* (2018) Generic meal patterns identified by latent class analysis: insights from NANS (National Adult Nutrition Survey). *Nutrients* **10**, 310.
19. Nurius PS & Macy RJ (2008) Heterogeneity among violence-exposed women: applying person-oriented research methods. *J Interpers Violence* **23**, 389–415.
20. von Eye A & Bergman LR (2003) Research strategies in developmental psychopathology: dimensional identity and the person-oriented approach. *Dev Psychopathol* **15**, 553–580.
21. Lanza ST & Rhoades BL (2013) Latent class analysis: an alternative perspective on subgroup analysis in prevention and treatment. *Prev Sci* **14**, 157–168.
22. Leech RM, Worsley A, Timperio A *et al.* (2017) Temporal eating patterns: a latent class analysis approach. *Int J Behav Nutr Phys Act* **14**, 3.
23. Miettunen J, Nordstrom T, Kaakinen M *et al.* (2016) Latent variable mixture modeling in psychiatric research – a review and application. *Psychol Med* **46**, 457–467.
24. DiStefano C & Kamphaus RW (2006) Investigating subtypes of child development: a comparison of cluster analysis and latent class cluster analysis in typology creation. *Educ Psychol Meas* **66**, 778–794.
25. Lu S, Qian Y, Huang X *et al.* (2017) The association of dietary pattern and breast cancer in Jiangsu, China: a population-based case-control study. *PLoS One* **12**, e0184453.
26. Zhao W, Hasegawa K & Chen J (2002) The use of food-frequency questionnaires for various purposes in China. *Public Health Nutr* **5**, 829–833.
27. Armstrong T & Bull F (2006) Development of the World Health Organization Global Physical Activity Questionnaire (GPAQ). *J Public Health* **14**, 66–70.
28. Fahey MT, Thane CW, Bramwell GD *et al.* (2007) Conditional Gaussian Mixture modelling for dietary pattern analysis. *J Royal Stat Society Ser A (Stat Soc)* **170**, 149–166.
29. Wedel M & Kamakura WA (2002) *Mixture Regression Models*. New York: Cambridge University Press.
30. Van Buuren S, Brand JP, Groothuis-Oudshoorn CG *et al.* (2006) Fully conditional specification in multivariate imputation. *J Stat Comput Simul* **76**, 1049–1064.
31. Santos RO, Gorgulho BM, Castro MA *et al.* (2019) Principal Component Analysis and Factor Analysis: differences and similarities in Nutritional Epidemiology application. *Rev Bras Epidemiol* **22**, e190041.
32. Varraso R, Garcia-Aymerich J, Monier F *et al.* (2012) Assessment of dietary patterns in nutritional epidemiology: principal component analysis compared with confirmatory factor analysis. *Am J Clin Nutr* **96**, 1079–1092.
33. MacCallum RC, Zhang S, Preacher KJ *et al.* (2002) On the practice of dichotomization of quantitative variables. *Psychol Methods* **7**, 19–40.
34. Rabe-Hesketh S & Skrondal A (2008) Classical latent variable models for medical research. *Stat Methods Med Res* **17**, 5–32.
35. Butler LM, Wu AH, Wang R *et al.* (2010) A vegetable-fruit-soy dietary pattern protects against breast cancer among postmenopausal Singapore Chinese women. *Am J Clin Nutr* **91**, 1013–1019.
36. Shin S, Saito E, Inoue M *et al.* (2016) Dietary pattern and breast cancer risk in Japanese women: the Japan Public Health Center-based Prospective Study (JPHC Study). *Br J Nutr* **115**, 1769–1779.
37. Kojima R, Okada E, Ukawa S *et al.* (2017) Dietary patterns and breast cancer risk in a prospective Japanese study. *Breast Cancer* **24**, 152–160.
38. Zhang CX, Ho SC, Fu JH *et al.* (2011) Dietary patterns and breast cancer risk among Chinese women. *Cancer Causes Control* **22**, 115–124.
39. Cui X, Dai Q, Tseng M *et al.* (2007) Dietary patterns and breast cancer risk in the shanghai breast cancer study. *Cancer Epidemiol Biomarkers Prev* **16**, 1443–1448.
40. Strom A, Hartman J, Foster JS *et al.* (2004) Estrogen receptor beta inhibits 17beta-estradiol-stimulated proliferation of the breast cancer cell line T47D. *Proc Natl Acad Sci USA* **101**, 1566–1571.
41. Gong L, Li Y, Nedeljkovic-Kurepa A *et al.* (2003) Inactivation of NF-kappa B by genistein is mediated via Akt signaling pathway in breast cancer cells. *Oncogene* **22**, 4702–4709.
42. Wu AH, Yu MC, Tseng CC *et al.* (2009) Dietary patterns and breast cancer risk in Asian American women. *Am J Clin Nutr* **89**, 1145–1154.
43. Keum N, Greenwood DC, Lee DH *et al.* (2015) Adult weight gain and adiposity-related cancers: a dose-response meta-analysis of prospective observational studies. *J Natl Cancer Inst* 107, djv088.
44. Nelson LR & Bulun SE (2001) Estrogen production and action. *J Am Acad Dermatol* **45**, S116–S124.
45. van den Brandt PA & Schulpen M (2017) Mediterranean diet adherence and risk of postmenopausal breast cancer: results of a cohort study and meta-analysis. *Int J Cancer* **140**, 2220–2231.
46. Castello A, Boldo E, Perez-Gomez B *et al.* (2017) Adherence to the Western, Prudent and Mediterranean dietary patterns and breast cancer risk: MCC-Spain study. *Maturitas* **103**, 8–15.
47. Harris HR, Willett WC, Vaidya RL *et al.* (2017) An adolescent and early adulthood dietary pattern associated with inflammation and the incidence of breast cancer. *Cancer Res* **77**, 1179–1187.

https://doi.org/10.1017/S1368980020004826 Published online by Cambridge University Press