

# Disentangling women's responses on complex dietary intake patterns from an Indian cross-sectional survey: a latent class analysis

Sabu S Padmadas<sup>1,\*</sup>, José G Dias<sup>2</sup> and Frans J Willekens<sup>3</sup>

<sup>1</sup>Division of Social Statistics & Southampton Statistical Sciences Research Institute, University of Southampton, Highfield, Southampton SO17 1BJ, UK; <sup>2</sup>Department of Quantitative Methods, Instituto Superior de Ciências do Trabalho e da Empresa (ISCTE), Lisbon, Portugal; <sup>3</sup>Netherlands Interdisciplinary Demographic Institute, The Hague, The Netherlands and Population Research Centre, University of Groningen, The Netherlands

Submitted 16 August 2004; Accepted 17 June 2005

## Abstract

*Objective:* To investigate the degree of individual heterogeneity related to complex dietary behaviour and to further examine the associations of different dietary compositions with selected characteristics.

*Design:* Latent class analysis was applied to data from the recent cross-sectional National Family Health Survey that collected information on the intake frequency of selected foods. Different responses regarding intake frequency were condensed into a set of five meaningful latent clusters representing different dietary patterns and these clusters were then labelled based on the reported degree of diet mixing.

*Setting:* Indian states.

*Subjects:* In total, 90 180 women aged 15–49 years.

*Results:* Three clusters were predominantly non-vegetarian and two were vegetarian. A very high or high mixed-diet pattern was observed particularly in the southern and a few north-eastern states. Many women in the very high mixed-diet cluster consumed mostly non-green/leafy vegetables on a daily basis, and fruits and other non-vegetarian diet on a weekly basis. In contrast, those in the low mixed-diet cluster consumed more than three-fifths of the major vegetarian diet ingredients alone on a daily basis. The affluent group that represented the low mixed-diet cluster were primarily vegetarians and those who represented the very high mixed-diet cluster were mostly non-vegetarians. The significant interrelationships of different characteristics highlight not only socio-economic, spatial and cultural disparities related to dietary practices, but also the substantial heterogeneity in diet mixing behaviour.

*Conclusions:* The results of this study confirmed our hypothesis of heterogeneous dietary behaviour of Indian women and yielded useful policy-oriented results which might be difficult to establish otherwise.

**Keywords**  
Dietary patterns  
Women  
India  
Latent class model  
Unobserved heterogeneity  
National Family Health Survey

Individuals experience different ways of life and this complexity is reflected in their eating customs and dietary habits<sup>1</sup>. Rapidly changing ways of life, growing inequalities in income and resources distribution, and the widening gap between rich and poor are some of the key issues that make food intake analyses more complex and difficult to understand<sup>2–4</sup>. This proposition holds true universally and especially in a context like India, where the health, socio-economic and demographic inequalities are larger both among individuals and across regions.

In the last few decades, the major source of dietary information in India has been the surveys conducted by the public health directorates of different states, the results of which were then published by the National Institute of Nutrition<sup>5</sup>. The National Nutrition Monitoring Board, set

up in 1972 as an integral part of the National Institute of Nutrition, periodically collects data on dietary intake and nutritional status based on representative multi-clustered samples from 10 selected states from different regions of India<sup>6</sup>. Other sources are the District Nutrition Profiles Surveys conducted in 15 selected states by the Food and Nutrition Board and the quinquennial consumer per-capita expenditure surveys of the National Sample Survey Organization<sup>7</sup>. Although, over time, these surveys have provided reasonable aggregate data on food consumption, they are not made available at the national level which has restricted regional or state comparisons. The second round of the National Family Health Survey (NFHS-2), conducted in India during 1998–99, collected individual-level information from ever-married women

\*Corresponding author: Email ssp@socsci.soton.ac.uk

aged 15–49 years on their daily, weekly and occasional consumption of selected foods, together with a set of demographic and socio-economic variables<sup>8</sup>.

Women's dietary behaviour in India at the national level has neither received adequate attention nor been analysed systematically. The present paper reports a method of analysis of dietary information using individual-level data which combines information from several indicator variables in a convenient manner. We explore the hypothesis of individual heterogeneity related to complex diet mixing behaviour based on the intake frequency of different foods using latent class (LC) analysis, using data from the NFHS-2. Two specific research questions are addressed: (1) Does LC analysis provide better insights into modelling dietary behaviour? (2) How can we classify certain groups of respondents according to various dietary compositions and different demographic, socio-economic, spatial and cultural characteristics?

A national-level analysis of dietary patterns holds considerable importance in India both from scientific and policy viewpoints. The scientific perspective addresses the need for a better understanding of dietary habits and the role of diet-related risk factors associated with non-communicable chronic diseases in later life, and the policy perspective aims to shape dietary guidance and evaluations for a comprehensive food policy that could be integrated with the national population and reproductive health policies<sup>9–13</sup>.

## Materials and methods

### Data

NFHS-2 was conducted between November 1998 and March 1999. The survey covered a representative sample of more than 90 303 eligible women aged 15–49 years from 91 196 households in 26 states. Further details are available elsewhere<sup>8</sup>. NFHS-2 is the first survey of its kind to record women's dietary intake information at the national level. From a total sample of 90 303 women, 90 180 provided complete responses which are considered for the analysis. The analysis was carried out for six regions comprising 26 Indian states, which represents more than 99% of India's total population. The states included in the analysis have considerable demographic, social and cultural disparities.

NFHS-2 asked ever-married women 'How often do you yourself consume the following items: daily, weekly, occasionally or never?' This question was not asked in the previous round of the NFHS. In NFHS-2 women were probed regarding the consumption of specific food items: milk or curd, pulses or beans, green leafy vegetables, other vegetables, fruits, eggs, and chicken, meat or fish. The food items listed were mentioned to the respondents and coded according to the frequency of consumption. The four category responses, i.e. the frequency of intake, provided in NFHS-2 reflect the immediate past and current

dietary habits of the women. The analysis in the present paper focuses on the frequency of dietary intake as response variable along with other independent variables of interest. Unfortunately, this was the only question on diet that was included in the respondent's background section of the NFHS-2 questionnaire. The survey did not ask specifically about any cereal (wheat/rice/corn) intake; however, it is well known that wheat and rice consumption forms a major part of the daily food of the Indian population. The survey did not provide any information related to the quantity or level of food consumption over time, and such analysis is beyond the scope of this paper.

### Method

We used the LC approach to examine the differential clustering of individual dietary behaviour based on the frequency of intake. LC analysis groups food intake data into a meaningful set of latent classes representing different dietary intake patterns that explain the association patterns among the observed variables. It tests whether a discrete latent variable, specified as a set of mutually exclusive classes, accounts for observed covariation among manifest, categorical variables<sup>14,15</sup>. A binary-type LC analysis of vegetable consumption patterns among 1028 US women was attempted earlier by Patterson *et al.*<sup>16</sup>. In our approach individuals are grouped based on the data from polytomous indicators, thereby decomposing the sample into clusters<sup>17</sup>. The parameters of the models were estimated using maximum likelihood methods based on the Expectation Maximum (EM) algorithm<sup>18,19</sup>. Two types of parameters were distinguished: (1) the latent class probabilities that represent the proportion of women in the sample who fall into each class; and (2) the conditional probabilities of each dietary pattern within each latent class. The resultant dietary cluster memberships (dependent variable) were analysed using multinomial logistic regression models to examine their association with selected demographic, spatial, socio-economic and cultural characteristics.

### Model estimation

We fitted a model with a maximum of eight latent classes, using 20 runs with different starting values, in order to avoid the local maxima. It should be noted that it is difficult to accept more than eight classes because of the number of parameters involved in the model. The traditional approach to select the best models is to use likelihood ratio tests. None the less, for LC models, likelihood ratio tests are inappropriate because the null hypothesis is defined on the boundary of the parameter space and consequently the asymptotic properties of the maximum likelihood estimation are invalid. An inferential alternative would be to apply the parametric bootstrap to test sequential null hypotheses of, say,  $S$  clusters against the alternative of  $S + 1$  clusters. However, for the number of latent classes

considered in our analyses, the null hypotheses were always rejected ( $P < 0.001$ ). An alternative approach consists in the selection of  $S$  clusters using information statistics, the most popular being the Akaike Information Criteria (AIC), the Bayesian Information Criteria (BIC) and the Consistent Akaike Information Criteria (CAIC)<sup>20–22</sup>. For these heuristic criteria, smaller values mean more parsimonious models. BIC and CAIC have the advantage of being dimension-consistent, i.e. they point to the right model with probability of unity as the sample size increases. The results suggested that the best solution had at least eight classes ( $S \geq 8$ ) corresponding to at least 175 free parameters. The solution with more than five classes had only a marginal effect for the given sample-consistent criteria (results not shown). Therefore we considered five latent classes in the model, corresponding to 109 independent parameters, without losing much information and also accounting for interpretation reasons. The EM algorithm for this model was programmed in MATLAB 6.5

(The Maths Work Inc., Natick, MA, USA, 2002). For a clear description of LC methodology, readers are referred to Wedel and Kamkura<sup>23</sup> and Vermunt and Magidson<sup>24</sup>.

## Results

### *Reported dietary intake patterns in India: a descriptive overview*

The complex discrepancy of dietary intake patterns across different Indian regions is shown in Table 1. More than 85% of women in India reported to have consumed pulses/beans and green leafy vegetables at least once a week. Kerala is an exception, where only 55% of women reported consumption of green leafy vegetables at least once in a week. Intake of egg and meat products is relatively low in many states, particularly in the north and east where poverty levels are high in states such as Bihar<sup>25</sup>. Respondents from Kerala and Goa located in the coastal area reported high intake of chicken/meat/fish; fish

**Table 1** Percentage distribution of ever-married women classified by dietary intake at least once a week, India and states, 1998–99

State	Type of food						Chicken, meat, or fish	Number of women
	Milk or curd	Pulses or beans	Green leafy vegetables	Other vegetables	Fruits	Eggs		
India	55.0	87.8	85.2	93.1	33.0	27.8	31.9	90303
North								
Delhi	73.3	91.2	86.8	92.8	57.8	21.2	15.1	2477
Haryana	93.2	99.3	99.2	99.2	54.8	7.7	3.8	2908
Himachal Pradesh	87.0	99.1	94.3	98.8	71.7	14.7	6.2	3012
Jammu & Kashmir*	72.1	68.5	85.5	88.3	44.0	14.2	31.1	2744
Punjab	91.1	99.2	99.1	99.5	50.7	10.8	3.6	2796
Rajasthan	70.7	81.4	77.8	78.9	20.5	6.1	7.8	6813
Central								
Madhya Pradesh	32.5	79.9	80.9	86.1	22.7	11.7	11.2	6941
Uttar Pradesh	57.2	88.0	90.0	90.7	19.0	9.9	8.7	9292
East								
Bihar	46.7	88.7	96.0	96.1	18.3	22.1	21.5	7024
Orissa	20.7	80.7	90.9	95.8	14.4	15.6	28.2	4425
West Bengal	25.0	76.3	91.4	98.7	15.0	43.5	69.0	4408
Northeast								
Arunachal Pradesh	19.9	51.2	95.6	72.7	28.9	33.5	57.4	1117
Assam	41.7	85.3	87.6	94.9	33.3	58.4	57.7	3441
Manipur	15.3	37.3	96.9	93.2	34.3	14.8	47.4	1435
Meghalaya	23.7	61.5	88.9	91.8	40.3	32.6	61.8	945
Mizoram	22.9	64.5	99.2	87.1	61.6	42.5	59.3	1048
Nagaland	82.7	59.6	96.3	80.6	40.9	30.2	72.3	818
Sikkim	72.4	82.9	94.9	87.5	28.8	26.8	57.1	1107
Tripura†	51.0	86.1	91.2	91.5	39.9	56.3	65.2	1104
West								
Goa	65.0	76.5	74.6	82.5	65.8	36.6	89.0	1246
Gujarat	80.0	97.0	74.1	99.2	44.4	14.0	12.4	3845
Maharashtra	47.3	94.5	87.9	91.1	44.7	34.4	38.2	5391
South								
Andhra Pradesh	72.0	92.3	72.7	95.7	47.6	59.7	56.7	4032
Karnataka	75.5	98.6	93.3	91.8	53.7	39.9	33.9	4374
Kerala	45.3	69.8	54.8	90.9	56.5	27.3	82.8	2884
Tamil Nadu	66.5	94.6	77.6	98.7	46.2	52.7	52.6	4676

Source: Second round of the National Family Health Survey (NFHS-2) (reference 8, p. 244). Less than 0.1% of cases were missing for all India and states. Weighted data were used for the analysis.

\* Jammu region of Jammu & Kashmir.

† At the time when the NFHS-2 report was published, the state of Tripura was not included because the fieldwork was not completed. We included Tripura in the analysis.

consumption particularly has been noteworthy in these states for a long time<sup>26,27</sup>. The overall consumption of fruits was also reported to be low, noticeably in the central and eastern regions<sup>27</sup>.

### **Description of the latent classes**

After ordering the identified classes, we labelled the clusters corresponding to diet mixing – the combination of various vegetarian and non-vegetarian dietary intakes. In the sample, 25.5% represented a very high mixed-diet cluster, 21.4% a high mixed-diet cluster, 20.6% a moderate mixed-diet cluster, and roughly 16% each a low and a very low mixed-diet cluster. The very high, high and moderate classes are predominantly non-vegetarian and the low and very low are vegetarian. The interpretations, although a little complex, reveal interesting dietary patterns (Fig. 1a–f).

The reported dietary intake of the high mixed-diet cluster indicates consumption of mostly vegetables other than green and leafy ones, and pulses/beans on a daily basis, whereas eggs and chicken/meat/fish are consumed on a weekly basis. Evaluation of responses of the low or very low mixed-diet cluster seems to indicate a vegetarian dietary pattern. For example, among women in the low mixed-diet cluster, more than three-fifths reported intake of milk/curd, pulses/beans and other vegetables on a daily basis whereas only negligible proportions seem to have consumed eggs and meat/chicken/fish. This clearly pinpoints the distinctiveness of the low mixed-diet (vegetarian) cluster. About 98% of women in this cluster reported intake of important vegetarian foods alone; more than 60% consumed vegetarian foods on a daily basis (Fig. 1e). The reported frequent fruit consumption on a daily basis is also relatively high in this cluster. About 98% in the low mixed-diet cluster appear never to include chicken/meat/fish in their diet, which probably indicates the segregation of a vegetarian group in the sample. Respondents in the moderate mixed-diet cluster seem to have consumed both vegetarian and non-vegetarian food although less frequently on a daily or weekly basis. The differences among the three non-vegetarian groups are, however, trivial.

### **Characterising dietary intake patterns**

Bivariate analyses were conducted to examine the association of selected individual and household characteristics with the clusters representing diverse dietary behaviour (results not shown separately). The spatial (place of residence, region), socio-economic and cultural profiles (standard of living, education, ethnicity, religion, occupation) of respondents differed considerably by different diet composition clusters. Little statistical variation in dietary compositions was observed in the case of selected demographic characteristics (respondent's age and current pregnancy status, presence of children below 5 years in the household); the differences seemed

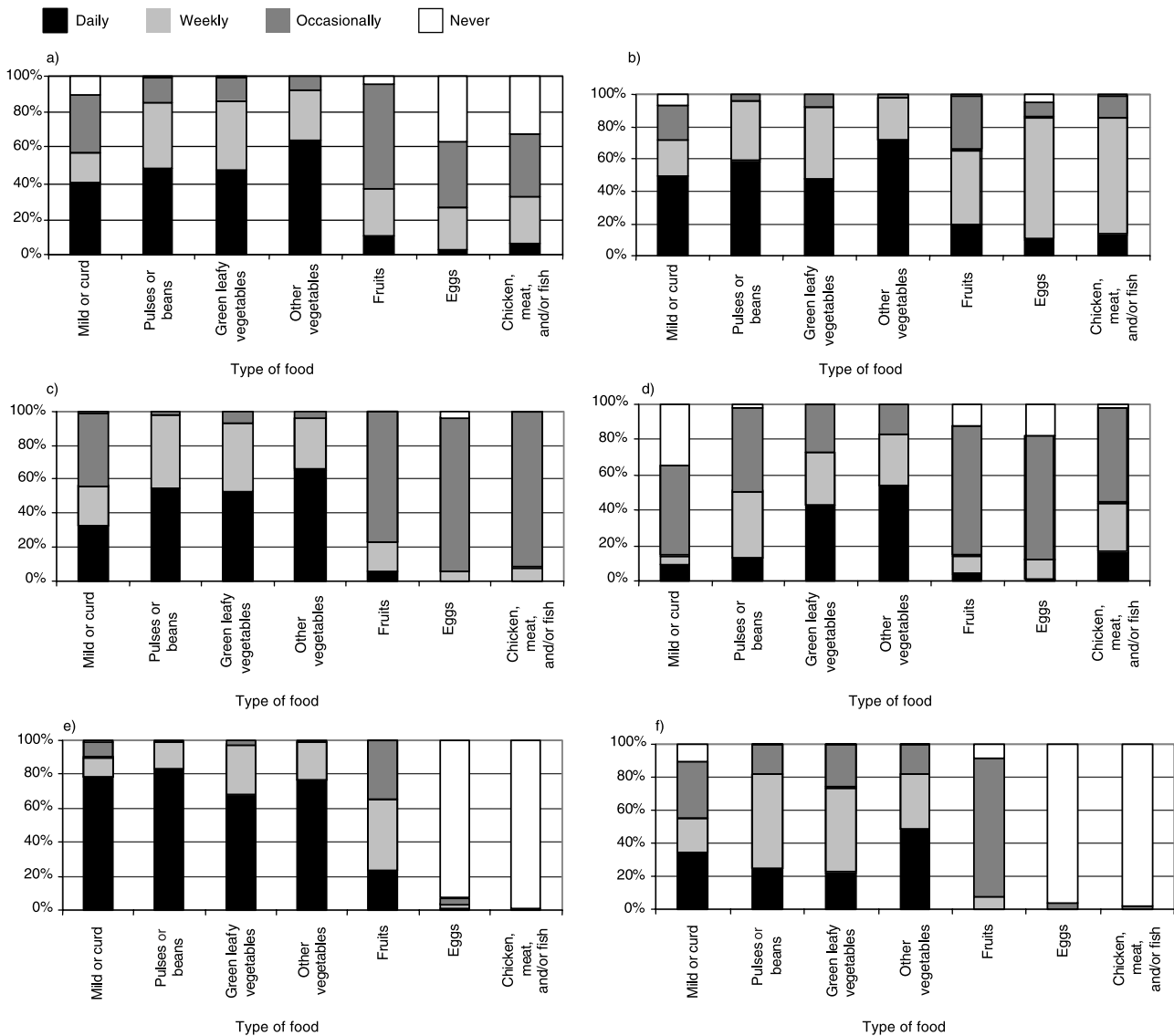
marginal with respect to aggregate figures. The differences observed across categories are summarised as follows.

Both very high and low mixed-diet clusters were predominantly urban respondents, whereas a significant proportion of rural respondents represented either the moderate or very low mixed-diet cluster. Regional variations indicated that considerable proportions of respondents from the southern regions were in the very high mixed-diet cluster whereas those from the northern regions represented the low mixed-diet cluster. Living standard differentials indicated that affluent respondents tended to fall in either the very high or low mixed-diet cluster. The standard of living score variable in the dataset is a composite index based on household utilities (type of household, toilet facilities, water and sanitation, landholdings) and consumer goods (e.g. television, radio, bicycle, car). Respondents without any schooling experience were more likely to report being in the very low mixed-diet (vegetarian) cluster, whereas those who had completed high school and above were more likely to be in the low mixed-diet or very high mixed-diet cluster. A significant proportion of Muslim respondents belonged to the very high mixed-diet cluster (non-vegetarian) whereas a very high proportion of Hindus were in the low or very low mixed-diet clusters (vegetarian). Respondents belonging to scheduled and other backward castes represented mostly the high mixed-diet cluster. Those engaged in agricultural activities fell mostly in the very low mixed-diet cluster. A few other variables were either less important or did not show any significant associations, for example respondent's current pregnancy status, total number of household members and respondent's current marital status.

### **Regression analysis**

The associations of selected characteristics with the individual responses reflected in different dietary classes were examined using multinomial logistic regression models (Table 2). The reference category of the dependent variable was respondents who represented the very low mixed-diet cluster. The models examined the spatial, socio-economic and cultural influences on women's dietary behaviour with a statistical control of selected demographic characteristics. For ease of interpretation, the results are presented as probabilities, expressed as percentages. The predicted probabilities represent an average woman, i.e. most representative in the sample, who is aged between 25 and 34 years, currently not pregnant, who has no children under 5 years, lives in a rural area under average conditions, non-working and without any schooling experiences, and without any defined ethnic background. The differences between categories of independent variables were statistically significant at  $P < 0.001$ , except respondent's current pregnancy status.

A very high or a high mixed-diet pattern was observed predominantly among urban women, whereas rural women represented mostly either the high or moderate



**Fig. 1** Dietary intake profiles (%) for the aggregate model (a) and the five latent classes: (b) very high mixed-diet cluster; (c) high mixed-diet cluster; (d) moderate mixed-diet cluster; (e) low mixed-diet cluster; (f) very low mixed-diet cluster

mixed-diet cluster. The share of the very low mixed-diet cluster was highest in the northern (5.9%) and central regions (4.1%). A very high mixed-diet pattern was seen mostly among respondents from the southern (55.8%), western (49.7%) and north-eastern region (47.9%), and observed least among those from the central region (18.3%). Women in affluent households were highly likely to fall in the very high mixed-diet cluster when compared with their poorer counterparts. Conversely, the share of the low mixed-diet pattern was significantly high among affluent groups (8.6%) compared with those in poorer backgrounds (2.3%). The results showed that women living in better conditions were almost equally as likely to be in either the very high or the low mixed-diet cluster. It is clear from the analysis that the affluent group who represented the low mixed-diet cluster were predominantly vegetarians and those who represented the very

high mixed-diet cluster were non-vegetarians. By the same token, educated women were found more likely to represent either the very high mixed-diet cluster or the low mixed-diet cluster than their counterparts. It should be noted that the descriptive analysis revealed a weak association between living conditions and women's education.

The very high or high mixed-diet pattern was characteristic of mostly Muslims (26.0%) and Christians (21.1%) compared with their Hindu or other religious counterparts (about 5.0%). The low mixed-diet cluster was mostly characteristic of other religious groups (55.0%) and the Hindus (40.8%). The share of the very high and high mixed-diet pattern was lowest among scheduled tribe (61.9%) compared with scheduled caste (72.7%) and other backward caste women (72.0%). A very low or low mixed-diet pattern was observed predominantly among

**Table 2** Adjusted predicted probabilities of diet mixing behaviour: results from multinomial logistic regression model ( $n = 90\,157$ )

Characteristic	Probability expressed as adjusted %				
	Very high mixed-diet	High mixed-diet	Moderate mixed-diet	Low mixed-diet	Very low mixed-diet
<i>Spatial</i>					
Place of residence					
Urban	40.1	33.1	20.0	3.1	3.7
Rural	26.0	40.1	23.4	4.6	5.9
Region					
North	26.0	40.1	23.4	4.6	5.9
Central	18.3	43.4	32.0	2.1	4.1
East	34.3	30.0	35.3	0.2	0.3
Northeast	47.9	19.2	32.6	0.1	0.1
West	49.7	21.7	23.8	2.1	2.7
South	55.8	19.6	23.8	0.6	0.2
<i>Socio-economic and cultural</i>					
Standard of living					
Low	18.7	41.2	32.5	2.3	5.3
Medium	26.0	40.1	23.4	4.6	5.9
High	35.2	35.4	15.4	8.6	5.4
Education					
High school or more	40.2	31.4	17.8	7.7	2.9
Secondary	36.2	33.4	21.0	5.7	3.6
Primary	31.6	35.2	23.7	4.8	4.7
None	26.0	40.1	23.4	4.6	5.9
Religion					
Others	5.3	13.9	6.3	55.0	19.5
Christian	20.1	37.1	28.8	4.3	9.7
Hindu	4.5	12.2	3.8	40.8	38.7
Muslim	26.0	40.1	23.4	4.6	5.9
Ethnicity					
Scheduled caste	28.2	44.5	23.1	1.6	2.6
Scheduled tribe	23.2	38.7	33.9	1.2	3.1
Other backward caste	25.8	46.2	19.4	3.1	5.6
None	26.0	40.1	23.4	4.6	5.9
Type of employment					
Professional	26.4	44.7	20.7	3.8	4.4
Services	26.1	37.2	29.4	3.1	4.2
Agriculture	20.6	39.3	30.0	2.7	7.3
Skilled manual	22.4	35.0	33.2	3.1	6.3
Unskilled manual	22.2	38.0	33.3	1.7	4.8
Not working	26.0	40.1	23.4	4.6	5.9
<i>Interaction</i>					
Living standards $\times$ residence*					
High $\times$ urban	50.1	28.9	11.9	5.9	3.2
Medium $\times$ urban	40.4	32.3	21.0	2.8	3.6
Low $\times$ urban	32.5	35.6	26.8	1.7	3.4
High $\times$ rural	37.2	33.7	16.0	8.1	5.1
Medium $\times$ rural	25.9	40.4	22.9	4.7	6.0
Low $\times$ rural	18.4	41.3	32.7	2.3	5.3

Other variables controlled for in the model but not shown in the table include current age of the respondent, children aged below 5 years in the household and respondent's current pregnant status. All variables controlled for in the model, except current pregnancy status, were statistically significant at  $P < 0.001$ .  $-2$  log-likelihood of the final model was 92 229, significant at  $P < 0.001$ .

\*Interaction effects were captured in a separate model with statistical control of selected spatial, demographic and cultural characteristics ( $-2$  log-likelihood: 92 168). The effects were statistically significant at  $P < 0.001$ .

non-working women or those working in the agricultural sector. Those working in the professional and services sectors were most likely to represent the very high mixed-diet cluster. Skilled or unskilled manual workers represented either the high or moderate mixed-diet cluster.

In a separate model, we examined the possible interaction effect between residence and living conditions on dietary behaviour after adjusting for potential

confounding demographic and other socio-economic and cultural variables (results are shown in Table 2 for convenience). The likelihood that an affluent urban woman represented the very high mixed-diet cluster was 50.1%, which decreased to 37.2% for her rural counterpart and to 18.4% for a woman who lived in rural areas under poor conditions. A low mixed-diet was mostly characterised by affluent women who lived in rural areas.

## Discussion

The foregoing analysis offers some insights into the dietary habits of the female population across Indian states and provides useful information about dietary behaviour that was hitherto unavailable nationally. The reported diet intake frequencies were pooled and then disaggregated into five classes using LC analysis. The five-cluster solution provided a good compromise between capturing the unobserved heterogeneity and disentangling the model complexity. These clusters were ordered based on the reported degree of diet mixing and were categorised as three predominantly non-vegetarian (very high, high and moderate) and vegetarian (low and very low) clusters.

The LC analysis employed in this study confirmed our hypothesis of heterogeneous dietary behaviour among different groups of Indian women at the individual level, and yielded useful results which might have been difficult to establish otherwise. Amongst women in the very high mixed-diet cluster, quite large proportions reported intake of non-green/leafy vegetables on a daily basis, and fruits and other non-vegetarian diet only on a weekly basis. This suggests three different possible scenarios. First, only a few households could afford to buy non-vegetarian foods on a daily basis. Second, although there is an income provision to afford non-vegetarian foods on a daily basis, sometimes it may be difficult to access such foods due to either a lack of production in certain regions (geographical constraints) or certain intra-household decisions on food consumption. Third, it may be either because of a lack of awareness about balanced nutritional intake or because of specific food preferences. On the other hand, 98% of respondents in the low mixed-diet group reported never having eaten chicken/meat/fish. Fruit consumption was generally very poor in all the clusters. Our data investigations showed that women who lived in poor conditions were the least likely to have consumed fruit compared with their counterparts. Community-level nutrition programmes could aim to increase women's awareness to include fruits in the diet at least on a weekly basis. In addition, efforts should focus on price subsidies so that poor people could have easy access to fruits.

The result that the low mixed-diet cluster consumes more than three-fifths of the major vegetarian diet ingredients on a daily basis is highly convincing. This particular group seems to be segregated in the northern region especially in Punjab and Haryana. The affluent group that represented the low mixed-diet was primarily following a vegetarian diet, whereas those who represented the very high mixed-diet cluster seemed to be following a non-vegetarian diet. The respondents in the moderate cluster lagged behind other clusters in terms of diet frequency, particularly of milk/curd and pulses/beans. The present analysis confirmed that these respondents were disadvantaged in terms of social and economic background.

The regional differentials in diet mixing behaviour were more likely to be due to differences in the distribution of religious and ethnic groups rather than socio-economic *per se*. Usually, the traditional orthodox Hindu Brahmin communities, especially in northern India, consume mainly vegetarian foods, unlike other religious groups<sup>28</sup>. These interlinked factors highlight not only the spatial, socio-economic and cultural disparities related to dietary practices but also the heterogeneity of dietary behaviour. Seasonal variations in agricultural production and supply, attitudes to consumption of healthy foods, and other cultural restrictions could influence dietary behaviour. The results suggest undertaking a more in-depth field investigation of dietary behaviour targeting particular socio-economic groups (e.g. certain religious and ethnic groups). Public health nutritionists could provide guidelines to improving national nutrition policies by emphasising the optimal dietary requirements of specific (poor) populations who are particularly disadvantaged in terms of diet and health status (undernourished). The importance of optimal or balanced diet mixing is reflected in the fact that the nutritional quality of the diet does improve with the consumption of a diet of greater diversity<sup>29,30</sup>.

Our findings recommend the need for a comprehensive and effective food policy in India to be integrated along with the national population and health policy. Although the country has succeeded in controlling population growth to a certain extent, the nutritional health of its people – especially women and children – remains a distant goal especially in light of the emerging coexistence of both obesity and undernutrition<sup>31</sup>. Our study reveals the need for a detailed demographic investigation of dietary intake between vegetarians and non-vegetarians both at the individual and population levels.

We underline here that the results of this study are only possible indications and might not reflect the actual diet mixing attitudes. Unfortunately, we could not explicitly differentiate between vegetarians and non-vegetarians from the NFHS-2 sample although we partly succeeded in differentiating various dietary compositions. An important data limitation of our study is the self-reported information available only for women, which was gathered at one point in time, i.e. cross-sectional and not longitudinal. Besides, we could not consider many other important nutrient-related foods (carbohydrates) including cereals in the model owing to lack of information. Therefore, the results presented herein should be treated with caution. Finally, to the best of our knowledge, this is one of the first attempts to model complex food intake patterns in developing countries using national-level individual data. Further extensions of the model for different populations and refinements of data collection methodology are suggested for a deeper understanding of dietary behaviour.

## Acknowledgements

We thank Saseendran Pallikadavath and John Micklewright for providing us with useful comments on an earlier version of this paper. Also, we gratefully acknowledge the invaluable editorial remarks and comments from anonymous reviewers which helped improve this manuscript substantially.

## References

- 1 Woods SC, Schwartz MW, Baskin DG, Seeley RJ. Food intake and the regulation of body weight. *Annual Review of Psychology* 2000; **51**: 255–77.
- 2 Kakwani N. *On Specifying Poverty Lines. Asia and Pacific Forum on Poverty*. Manila: Asian Development Bank, 2001.
- 3 Popkin BM. The shift in stages of the nutrition transition in the developing world differs from past experiences. *Public Health Nutrition* 2002; **5**: 205–14.
- 4 Shetty PS. Nutrition transition in India. *Public Health Nutrition* 2002; **5**: 175–82.
- 5 Srinikantia SG. *The National Nutrition Monitoring Bureau*. Hyderabad: Nutrition Foundation of India, 1998.
- 6 National Institute of Nutrition. *Nutrition in India: UN/ACC/SCN Country Case Study Supported by UNICEF*. Hyderabad: National Institute of Nutrition, 1992.
- 7 Shariff A, Mallick AC. Dynamics of food intake and nutrition by expenditure class in India. *Economic and Political Weekly* 1999; **34**: 1790–800.
- 8 International Institute for Population Sciences (IIPS). *National Family Health Survey (NFHS-2), 1998–99*. Mumbai/Washington, DC: IIPS/ORC Macro, 2000.
- 9 Reddy KS. Cardiovascular diseases in the developing countries: dimensions, determinants, dynamics and directions for public health action. *Public Health Nutrition* 2002; **5**: 231–7.
- 10 Hu FB. Diet, lifestyle, and risk of type 2 diabetes mellitus in women. *New England Journal of Medicine* 2001; **345**: 790–7.
- 11 Gopalan C. Diet related non-communicable diseases in South and South East Asia. In: Shetty PS, McPherson K, eds. *Diet, Nutrition and Chronic Disease: Lessons from Contrasting Worlds*. London: John Wiley & Sons, 1997; 10–23.
- 12 Chadha SL, Gopinath N, Shekawat S. Urban–rural differences in the prevalence of coronary heart disease and its risk factor in Delhi. *Bulletin of the World Health Organization* 1997; **75**: 31–8.
- 13 Kant AK, Schatzkin A, Ziegler RG. Dietary diversity and subsequent cause-specific mortality in the NHANES I epidemiological follow-up study. *Journal of the American College of Nutrition* 1995; **14**: 233–8.
- 14 Clogg CC. Latent class models. In: Arminger G, Clogg CC, Sobel ME, eds. *Handbook of Statistical Modelling for the Social and Behavioral Sciences, Volume 14*. New York: Plenum, 1995; 233–8.
- 15 Goodman LA. Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika* 1974; **61**: 215–31.
- 16 Patterson BH, Dayton CM, Graubard BI. Latent class analysis of complex survey data: application to dietary data. *Journal of the American Statistical Association* 2002; **97**: 721–9.
- 17 Lazarsfeld PF, Henry NW. *Latent Structure Analysis*. New York: Houghton Mifflin, 1968.
- 18 Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society* 1977; **B39**: 1–38.
- 19 McLachlan GJ, Peel D. *Finite Mixture Models*. New York: John Wiley & Sons, 2000.
- 20 Akaike H. A new look at statistical model identification. *IEEE Transactions on Automation and Control* 1974; **19**: 716–23.
- 21 Schwarz G. Estimating the dimension of a model. *Annals of Statistics* 1978; **6**: 461–4.
- 22 Bozdogan H. Model selection and Akaike's information criterion (AIC): the general theory and its analytical extensions. *Psychometrika* 1987; **52**: 345–70.
- 23 Wedel M, Kamkura W. *Market Segmentation: Conceptual and Methodological Foundations*. Boston, MA: Kluwer Academic Publishers, 1998.
- 24 Vermunt JK, Magidson J. Latent class models for classification. *Computational Statistics and Data Analysis* 2003; **41**: 531–7.
- 25 Dréze J, Sen A. *The Political Economy of Hunger. Vol. III. Endemic Hunger*. Oxford: Oxford University Press, 1991.
- 26 Achaya KT. *Fat Intakes in India*. Hyderabad: Nutrition Foundation of India, 1986.
- 27 Food and Agricultural Organization of the United Nations (FAO). *FAO Nutrition Country Profile of India*. Rome: FAO, 1998.
- 28 Sinha R, Anderson DE, McDonald SS, Greenwald P. Cancer risk and diet in India. *Journal of Postgraduate Medicine* 2003; **49**: 222–8.
- 29 Hatloy A, Torheim LE, Oshaug A. Food variety – a good indicator of nutritional adequacy of the diet? A case study from an urban area in Mali, West Africa. *European Journal of Clinical Nutrition* 1998; **52**: 891–8.
- 30 Drewnowski A, Henderson SA, Shore AB, Fischler C, Preziosi P, Hercberg S. Diet quality and dietary diversity in France: implications for the French paradox. *Journal of the American Dietetic Association* 1996; **96**: 663–9.
- 31 Griffiths PL, Bentley ME. The nutrition transition is underway in India. *Journal of Nutrition* 2001; **131**: 2692–700.