Public Health Nutrition: 8(2), 159–169 DOI: 10.1079/PHN2004670

Who eats four or more servings of fruit and vegetables per day? Multivariate classification tree analysis of data from the 1998 Survey of Lifestyle, Attitudes and Nutrition in the Republic of Ireland

Sharon Friel[1,*], John Newell[2] and Cecily Kelleher[3]

[1]Centre for Health Promotion Studies, National University of Ireland, Block T, Distillery Road, Galway, Republic of Ireland: [2]Department of Mathematics, National University of Ireland, Galway, Republic of Ireland: [3]Department of Epidemiology and Public Health Medicine, University College Dublin, Republic of Ireland

Submitted 20 April 2004: Accepted 11 August 2004

Abstract

Objective: To identify, using the novel application of multivariate classification trees, the socio-economic, sociodemographic and health-related lifestyle behaviour profile of adults who comply with the recommended 4 or more servings per day of fruit and vegetables.

Design: Cross-sectional 1998 Survey of Lifestyle, Attitudes and Nutrition.

Setting: Community-dwelling adults aged 18 years and over on the Republic of Ireland electoral register.

Subjects: Six thousand five hundred and thirty-nine (response rate 62%) adults responded to a self-administered postal questionnaire, including a semi-quantitative food-frequency questionnaire.

Results: The most important determining factor of compliance with the fruit and vegetable dietary recommendations was gender. A complex constellation of sociodemographic and socio-economic factors emerged for males whereas the important predictors of 4 or more servings of fruit and vegetable consumption among females were strongly socio-economic in nature. A separate algorithm was run to investigate the importance of health-related lifestyle and other dietary factors on compliance with the fruit and vegetable recommendations. Following an initial split on compliance with dairy recommendations, a combination of non-dietary behaviours showed a consistent pattern of healthier options more likely to lead to compliance with fruit and vegetable recommendations. There did, however, appear to be a compensatory element between the variables, particularly around smoking, suggesting the non-existence of an exclusive lifestyle for health risk.

Conclusions: Material and structural influences matter very much for females in respect to compliance with fruit and vegetable recommendations. For males, while these factors are important they appear to be mediated through other more socially contextual-type factors. Recognition of the role that each of these factors plays in influencing dietary habits of men and women has implications for the manner in which dietary strategies and policies are developed and implemented.

Keywords
Multivariate analysis
Classification trees
Diet
Social variation

Population-based studies world-wide have repeatedly shown that not only is dietary intake strongly age- and sex-related[1], but also that people from socially advantaged positions consume more nutritionally balanced diets in line with dietary recommendations, progressively becoming more unbalanced with decreasing socio-economic status[2–9]. Dietary behaviour is also influenced by social support-type factors such as marital status. Healthier diets, specifically fruit and vegetable consumption, are observed more commonly among persons who are married or living with someone, especially so for males[7,10].

Analytical interpretation of dietary data is driven by the survey measures and statistical methods used[11]. Traditionally there have been three main types of analyses of dietary data: examination of dietary compliance, investigation of possible associations between diet and disease and evaluation of nutrition education programmes[12]. Nutrition health education and promotion is based on recommended dietary allowances that are often represented visually by food pyramids, plates or circles[13–15], designed to give a general sense of the relative proportions and frequency of servings of foods and food

*Corresponding author: Email sharon.friel@nuigalway.ie

groups that contribute to a healthy diet among the general population.

Food choice is not an isolated action but rather is carried out and influenced by the economic, social and environmental context. The resulting large number of potential explanatory variables for dietary choice presents several analytical problems. First, it is not always clear what technique should be used to identify the best subset of explanatory variables out of all those available and, second, there is a strong argument for using as parsimonious a set of explanatory variables as possible. In recent years a number of studies have analysed food intakes and other lifestyle factors in such a way as to look at clusters of risk factors which may introduce a health risk that is greater than the risk expected from the sum of the individual risk factors[16–19]. Such an approach may allow discrimination of nutritional risk and hence facilitate targeted intervention[20]; however, within such statistical techniques there may be problems of collinearity[21].

Data mining is a term referring to the process of discovering patterns and structure in data. One such non-parametric approach, typically suited to both categorical and continuous classification problems with many potential explanatory variables, is the classification and regression tree (CART)[22]. This approach provides an alternative to predictive analyses performed using standard regression modelling methods such as logistic regression and classical classification methods such as stepwise discriminant analysis, which may be used to classify the characteristics of respondents who exhibit healthy versus unhealthy dietary behaviours. Tree-based approaches have an advantage over such classical methods because they do not have to conform to the same distribution restrictions (there is no assumption of a linear model) and are particularly useful when predictors may be associated in some non-linear or non-additive fashion[22–24]. Classification trees are so called because the primary method of display is in the form of a binary tree. They are used to predict class membership of a categorical dependent variable from their measurements on one or more predictor variables. A search is carried out across all the predictor variables to determine the best univariate split of a predictor variable that generates the greatest improvement in predictive accuracy. In doing so, a better understanding is gained of the structure of the relationships of the predictor variables (e.g. education, smoking status) and their combinations or interactions in terms of dependent variable class membership prediction (e.g. consumption of recommended daily servings of fruit and vegetables or not).

The present paper aims to describe the social variation in Irish adults' compliance with fruit and vegetable recommendations using data from the 1998 Survey of Lifestyle, Attitudes and Nutrition (SLAN). A multivariate classification tree approach is used in a novel way to determine the social status and non-diet-related behaviour profile of those who comply with fruit and vegetable dietary recommendations, and in doing so identify the sociodemographic, socio-economic factors and health-related lifestyle behaviours most predictive of compliance with these recommendations.

Methods and materials

Subjects

A cross-sectional survey, using a stratified probability sampling design, was undertaken in spring 1998. A two-stage random sample was drawn based on the adult population in each of the Republic of Ireland's 26 counties and was proportionately distributed according to the urban/rural breakdown in each county. Within each county the sampling unit was the district electoral division (DED) and the required number of urban and rural DEDs was ascertained based on 1996 Census data. Within each DED a random sample of 50 Irish adults aged 18 years and over on the electoral register was generated by a subsidiary company of the national postal system. Each selected adult was sent a self-administered questionnaire, plus explanatory letters and prepaid reply envelopes, of which 6539 (62%) were returned.

Instrument

Dietary assessment

A detailed description of the semi-quantitative food-frequency questionnaire (SQFFQ) used in the 1998 SLAN is presented elsewhere[25] but briefly the survey instrument comprised eight sections and included a 149 food item SQFFQ for the purposes of estimating usual food and nutrient intakes. An adapted version of the validated SQFFQ used in the British arm of the European Prospective Investigation of Cancer (EPIC) study[26,27] was utilised. The food items represented the whole diet and included 11 fruit and 27 vegetable items. Subjects were asked to indicate their average use of each food item over the last year. The frequency categories offered were 'never or less than once per month', 1–3 per month', 'once a week', '2–4 per week', '5–6 per week', 'once a day', '2–3 per day', '4–5 per day' and '6+per day'.

Non-diet-related lifestyle behaviours

A range of lifestyle-related indicators was recorded in the questionnaire. Questions from previously executed national and international population-based surveys were utilised to enhance comparability and reliability[28]. Those measures incorporated in this analysis include smoking, alcohol, physical exercise and self-reported weight and height. Respondents were asked if they were current occasional or regular cigarette smokers. Alcohol consumption was described by asking if the respondent drank alcohol, and if so in a typical week how many days they drank and how many drinks they had on average at

https://doi.org/10.1079/PHN2004670 Published online by Cambridge University Press

each drinking occasion. The number of units of alcohol consumed was derived by multiplying the number of days on which alcohol was consumed by the number of drinks consumed at the time. For this analysis, units of consumption are dichotomised into exceeding the recommended sensible weekly units (21 for males, 14 for females) or not. Respondents were asked to indicate how many times in a typical 7-day period they participated in strenuous, moderate and mild exercise, with examples provided of each level of activity. In this paper, responses are dichotomised into participation in any physical activity or not. Respondents were also asked to report their height and weight and from this their body mass index (BMI; weight in kg divided by the square of height in m) was calculated.

Social status indicators

A range of socio-economic and sociodemographic indicators was included in the SLAN questionnaire. This analysis uses information pertaining to the respondents' age, sex, education level, employment status, household tenure, marital status, number of people living in the household and locality of dwelling (rural/urban). The occupation of the principal wage earner in the household was also obtained and used to categorise respondents into social classes 1–6 based on the Irish Census 1996 classification system[29]. Social classes 1–2 correspond to professional, managerial and technical; social classes 3–4 include non-manual and skilled-manual occupations; and social classes 5–6 relate to semi-skilled and unskilled labour. Health care, including primary health practitioner services, is provided to all below an arbitrary level of income through the General Medical Services Scheme (GMS). Eligibility for the GMS, also known as medical card eligibility, is assessed at regional health authority level and factors like age, income and post-retirement means are taken into account.

Data analysis

Prior to the performance of any statistical analysis, excessive missing responses and outliers were identified in the data. Initially, questionnaires were excluded from the overall dataset if the entire food frequency section or two or more pages of the food frequency section were left blank[30,31]. Following removal of 413 blank questionnaires, the data were checked for outliers and a further 147 respondents were excluded based on their estimated energy intake being ± 2.5 standard deviations from the sample mean[32]. A final sample of 5979 was used in all subsequent analyses in this paper.

Within Ireland, dietary recommendations are graphically represented by a food pyramid consisting of five shelves. There is a recommended number of servings per day from each shelf that together combine to give an overall healthy and nutritious diet for the general adult population[13,33]. The bottom shelf relates to cereals, breads and potatoes, of which 6 or more servings per day are recommended. Four or more servings per day of fruit and vegetables are indicated on the next shelf. The third and fourth shelves correspond to dairy products including milk and cheese (3 servings per day recommended) and meat, fish or poultry (2 servings per day recommended). At the top of the pyramid there are mainly processed foods, high in saturated fats and simple sugars. It is suggested that these foods are eaten sparingly but is not more specific. For the purposes of this work, based on advice from the consultant dietitian for the Ministry for Health and Children, up to 3 servings per day from the top shelf has been used as the notional value. In order to estimate the number of servings from each food shelf of the pyramid consumed on a daily basis, each frequency option was scaled relative to a single daily serving. The number of servings from a shelf was then calculated by adding the scaled value for each relevant food item.

Chi-square tests were used to test for differences in the proportionate response of those consuming the recommended number of servings from each shelf of the food pyramid, stratified by age, across each social category. Similarly, differences in other diet and health-related lifestyle behaviours were also tested by age and sex using the chi-square test. Analyses were performed using SPSS software (SPSS Inc., Chicago, IL, USA) and the statistical significance level was set at 1% due to the large number of tests being performed.

Clearly, interactions exist between social status indicators and also between other lifestyle behaviours which may affect food choice and may be investigated formally using multivariate techniques[2,10]. *A priori* it was decided that fruit and vegetable consumption was strongly related to socio-economic circumstances, a finding noted in much of the international literature. It was decided to concentrate on this particular food shelf and to determine, using classification trees, both the social status and the lifestyle behaviour profile of respondents who did and did not consume the recommended 4 or more servings per day of fruit and vegetables. The classification tree derived in this analysis employed the CART approach, which involved successive binary partitioning of the dataset by identifying, at each partition step, which predictor variable best separates out the remaining observations in terms of class membership (i.e. compliance with fruit and vegetable recommendations or not). The purpose of the classification tree is to determine which social status and lifestyle behaviours are the strongest predictors of compliance with the fruit and vegetable dietary recommendations. A disadvantage of a tree-based approach is that it is not founded on a probabilistic model, and consequently there is no probability or confidence interval associated with predictions derived from using the tree to classify a new set of data. The only 'confidence' available in the accuracy of the results produced is based on the tree's predictive accuracy using cross-validation on

https://doi.org/10.1079/PHN2004670 Published online by Cambridge University Press

the original data. Such cross-validation, however, provides an alternative approach to estimate the relative predictive power of a model[34].

Two independent trees were developed using S-Plus statistical software (Insightful Corp., Reinach, Switzerland), both aiming to determine the profile and strength of each predicting factor for those respondents consuming 4 or more servings of fruit and vegetables per day. One tree was based on the sociodemographic and socio-economic variables (Table 1) and the other on health-related lifestyle behaviours (Table 2). It was initially intended to combine all variables into one tree; however, the complexity of the partitioning made it impossible to interpret the data. It is recognised that using two separate trees will not fully capture all interactions between independent variables. In both tree analyses the dependent variable is compliance with the recommended 4 or more servings per day of fruit and vegetables: no means <4 servings of fruit and vegetables, yes means ≥ 4.

A particular predictor variable enters the calculation only when it is required at a particular decision node on the basis of a splitting rule, and only one variable is used at each decision node. The predictive accuracy is measured by comparing the relative homogeneity of cases at each split; for classification tree problems the Gini index of node impurity is the measure most commonly chosen[22]. In this analysis the Gini criterion was used and did not incorporate prior information (such as the probability *a priori* of being in either category of the response variable). Data dropping 'down' a tree encounters one binary decision at a time until a terminal node is reached. Each terminal node was set to require a minimum node size of 50 individuals and tree pruning was based on 10-fold cross-validation in order to determine the 'best' tree.

Table 1 Sociodemographic and socio-economic tree: independent candidate variables

Variable	Categories
Gender	0 = male, 1 = female
Age group (years)	0 = 18–34, 1 = 35–64, 2 = 65+
Social class group	0 = SC 1/2, 1 = SC 3/4, 2 = SC 5/6
Highest attained level of education	0 = tertiary, 1 = secondary, 2 = none/primary
Employment status	0 = employed, 1 = other
Medical card status	0 = no medical card, 1 = yes
Household tenure	0 = owned with mortgage/outright, 1 = rented/other
Marital status	0 = married, 1 = single, 2 = previously
Locality of dwelling	0 = rural, 1 = urban
Number living in household	0 = living with others, 1 = alone

Results

Table 3 summarises the main social status characteristics of respondents. Comparison of the data with the most recent national census figures[29] shows similar age, gender and social class distributions, although females in lower social classes are slightly under-represented.

Bivariate analyses identified the association between individual social status factors, lifestyle behaviours and dietary choices, in particular compliance with the recommended 4 or more daily servings from the fruit and vegetable shelf. As seen in Table 4, statistically significant differences in compliance with dietary recommendations existed predominantly across sexes, age groups, marital status, number of people living in the household, level of education, employment status, medical card eligibility, social class and household tenure.

There was significant age and gender variation in respondents' non-diet health-related lifestyle behaviours. Table 5 shows the response for each behaviour of males and females, broken down across three age groups. Significantly more females than males participated in moderate weekly exercise ($\chi^2 = 49.2$, $P < 0.01$) and had a normal BMI ($\chi^2 = 191.7$, $P < 0.01$). On the other hand, males were more likely to do strenuous physical activity ($\chi^2 = 94.9$, $P < 0.01$), smoke more cigarettes ($t = 5.87$, $P < 0.01$) and exceed the recommended weekly alcohol limits ($\chi^2 = 18.9$, $P < 0.01$). Forty per cent of males compared with 24% of females were overweight and 12% were obese as opposed to 8% of females. There was significant variation across age in the numbers of respondents, both male and female, partaking in the different lifestyle behaviours. Significantly greater numbers of respondents aged 65 years and over took no

Table 2 Health-related lifestyle behaviours tree: independent candidate variables

Variable	Categories
Regular cigarette smoking status	0 = non-smoker, 1 = smoker
Body mass index	1 = normal, 2 = overweight, 3 = obese
Participation in physical activity	0 = doing exercise, 1 = no exercise
Compliance with recommended number of weekly alcoholic drinks	0 = within limits, 1 = exceeding limits
Compliance with the recommended 6 or more servings from the cereals, breads and potatoes shelf of the food pyramid	0 = less than 6 servings, 1 = 6 or more servings
Compliance with the recommended 3 servings from the dairy shelf of the food pyramid	1 = less than 3 servings, 2 = 3 servings, 3 = more than 3 servings
Compliance with the recommended 2 servings from the meat, fish and poultry shelf of the food pyramid	1 = less than 2 servings, 2 = 2 servings, 3 = more than 2 servings
Consumption of less than 3 servings from the top shelf of the food pyramid	0 = less than 3 servings, 1 = 3 or more servings

https://doi.org/10.1079/PHN2004670 Published online by Cambridge University Press

Table 3 Sociodemographic and socio-economic characteristics of respondents included in dietary analysis. Values are expressed as *n* (%)

	Overall sample ($n = 5979$)	Males ($n = 2739$)	Females ($n = 3129$)
Age group (years)			
18–34	2158 (37.1)	980 (36.1)	1187 (38.4)
35–64	2751 (47.3)	1321 (48.7)	1421 (46.0)
65+	913 (15.7)	411 (15.2)	483 (15.6)
Social class			
SC 1/2	1700 (40.9)	709 (37.4)	987 (44.7)
SC 3/4	1612 (38.8)	732 (38.6)	858 (38.8)
SC 5/6	841 (20.3)	456 (24.0)	364 (16.5)
Level of education			
Tertiary	1665 (30.5)	782 (30.7)	890 (30.8)
Secondary	2682 (49.1)	1176 (46.2)	1493 (51.6)
None/primary	1111 (20.4)	589 (23.1)	509 (17.6)
Home tenure			
Owned outright/ with mortgage	4657 (80.6)	2175 (80.9)	2476 (80.5)
Rented/other	1118 (19.4)	514 (19.1)	599 (19.5)
Employment status			
Working	2932 (52.7)	1639 (63.9)	1268 (43.4)
Other	2629 (47.3)	926 (36.1)	1653 (56.6)
Medical card			
No	4039 (71.5)	1919 (74.2)	2085 (69.7)
Yes	1607 (28.5)	667 (25.8)	905 (30.3)
Marital status			
Married/cohabiting	3150 (54.3)	1428 (52.9)	1713 (55.5)
Previously married	651 (11.2)	192 (7.1)	447 (14.5)
Single/never married	1997 (34.4)	1077 (39.9)	928 (30.1)
Locality of dwelling			
Urban	2706 (47.9)	1267 (48.6)	1385 (46.6)
Rural	2939 (52.1)	1342 (51.4)	1587 (53.4)
Number in household			
Live alone	801 (13.8)	388 (14.6)	393 (12.9)
More than one person	4985 (86.2)	2273 (85.4)	2660 (87.1)

SC 1/2 – professional, managerial and technical; SC 3/4 – non-manual, skilled manual; SC 5/6 – semi-skilled and unskilled manual; previously married – widowed or divorced.

weekly exercise at all (males: $\chi^2 = 67.9, P < 0.01$; females: $\chi^2 = 224.2, P < 0.01$) and had higher levels of obesity.

Classification tree results for compliance with fruit and vegetable daily recommendations

Sociodemographic and socio-economic profile

All 5979 respondents were entered into the classification tree algorithm but only those remained for which responses were available in all variables. A best fit was obtained with 19 nodes and a misclassification error of 32.6%.

Results from the classification tree of the social status variables in terms of profile and predicting consumption of 4 or more servings of fruit and vegetables per day revealed quite different patterns for males and females (Fig. 1). The most important determining factor of compliance with the fruit and vegetable recommendations was gender. A complex constellation of determining factors then emerged for males whereas the important factors for predicting the consumption of 4 or more servings of fruit and vegetables among females were strongly socio-economic in nature. Among females, medical card status, education and social class remained as the most important predictors for compliance with the fruit and vegetable dietary recommendations. Females who did not hold a medical card were highly likely to comply with the recommendations. If in possession of a medical card but with tertiary or secondary level education, then the chances of consuming 4 or more servings of fruit and vegetables increased. Social class status then became important. Females with a medical card, of little or no formal education and in social classes 3 and 4, compared with other social groups, were more likely to comply with the fruit and vegetable recommendations.

For males, on the other hand, a combination of socio-economic and social support-type factors, i.e. social class, marital status, medical card status, employment status, education, age and home tenure, all remained predictive of fruit and vegetable recommendation compliance. However, the interactions between the factors were much more complex than anything observed for females. Social class status was the strongest divisor of males in terms of consuming 4 or more servings of fruit and vegetables per day. Males in social class groups 1 and 2 were likely to comply with the recommendations. If in the remaining social classes 3–6, marital status became important. For non-married (either single or previously married) males in these social classes, medical card ownership signified non-compliance with the fruit and vegetable recommendation. Non-married males in social classes 3–6 but with tertiary level education or such males with up to secondary level education and in employment were likely to consume 4 or more servings per day of fruit and vegetables. For married males in social classes 3–6, if not in employment their chances of complying with the recommendations increased. Employed men over the age of 65 years were not likely to consume 4 or more servings of fruit and vegetables. Social class status became important again for employed, married men under the age of 65 years. Education, age and medical card status were important in predicting compliance with the fruit and vegetable recommendation for those in social classes 3 and 4. Home tenure and education level became the main predictors of fruit and vegetable compliance for married, employed males aged less than 65 years in social classes 5 and 6.

Diet and health-related lifestyle behaviour profile

A separate algorithm was run to investigate the health-related lifestyle and other dietary factor profile of those complying or not with the fruit and vegetable recommendations. A best fit was obtained with 10 nodes, correctly classified 66% of individuals and showed an initial split on the dairy shelf of the food pyramid (Fig. 2).

Regardless of all other factors, those who consumed on a daily basis more than 3 servings of dairy foodstuffs were also likely to consume 4 or more servings of fruit and vegetables. Compliance with the fruit and vegetable

https://doi.org/10.1079/PHN2004670 Published online by Cambridge University Press

Table 4 Social status variations in percentages consuming the recommended number of daily servings from each shelf of the food pyramid

	Males (%)					Females (%)				
	CBP	FV	Dairy	MFP	Top	CBP	FV	Dairy	MFP	Top
Age group (years)										
18–34	42.8	51.7	24.5*	30.5*	8.6*	34.7*	67.4*	20.0	39.3*	11.4*
35–64	42.3	55.4	22.2	39.3	14.1	40.0	73.7	22.7	41.3	15.5
65+	45.8	53.8	22.3	48.9	19.7	40.7	60.3	23.2	38.4	19.1
Social class										
SC 1/2	44.1	63.7*	24.1	39.4*	10.0	34.1	76.7*	23.9*	42.1*	11.2
SC 3/4	41.0	52.7	23.8	32.3	11.8	39.5	69.7	21.4	43.5	13.6
SC 5/6	44.6	44.9	21.2	33.9	10.3	40.8	65.5	16.6	34.3	12.6
Level of education										
Tertiary	42.4	61.3*	24.5*	39.0*	9.8	34.4	76.8*	24.2*	43.6*	11.7*
Secondary	43.1	53.2	22.3	33.9	11.4	39.8	69.5	21.0	39.4	13.4
None/primary	45.9	44.9	22.5	37.1	20.2**	40.5	54.7	21.4	34.6	22.4
Employment status										
Employed	42.3	55.6	22.9	36.3	11.2	33.5	72.4*	23.5	42.1	14.2
Other	44.6	51.6	23.7*	36.4	15.9*	40.6*	66.4	20.2	38.2	14.5
Medical card										
No	43.2	57.6*	23.3*	37.0	11.8	37.0	74.0*	21.6	42.7*	13.5
Yes	42.4	43.0	23.8	34.1	16.6*	40.9	58.9	22.5	33.8	16.2
Household tenure										
Owned	44.4*	56.0*	22.4	37.3*	13.0	38.0	71.1*	22.1	41.8*	14.7
Other	36.5	44.7	25.2	30.2	12.4	38.8	62.8	20.4	33.4	12.9
Marital status										
Married/cohabiting	44.6	58.9*	22.4*	38.8	12.3	39.3	72.2*	22.0*	41.3*	13.2*
Single/never married	42.8	48.7	23.3	31.2	12.0	34.9	67.1	21.0	37.4	15.1
Previously married	34.3	44.0	23.7	43.2*	21.5*	41.8	62.1	22.7	40.5	18.9
Locality of dwelling										
Urban	37.9	53.9	23.0	34.7	13.6	33.3*	67.8	24.3*	41.5*	14.9
Rural	47.7*	53.0	23.3	37.5	12.9	42.4	70.4	19.9	38.2	14.5
Number in household										
Live alone	34.8	44.8	25.3*	41.6*	21.6*	34.7	63.9	18.5	41.3*	21.3*
> 1 person	44.5*	55.6*	22.6	35.1	11.1	38.7	70.2	22.2	40.1	13.3

Food pyramid shelves: CBP – cereals, bread & potatoes; FV – fruit & vegetables; Dairy – dairy and alternatives; MFP – meat, fish & poultry; Top – foods high in sugars and high in fats.
* Significant difference between categories: $P < 0.01$.

Table 5 Percentage responses of non-diet health-related lifestyle behaviours, by age and gender

		Males				Females			
	Overall ($n = 2735$)	18–34 years ($n = 980$)	35–64 years ($n = 1321$)	65 + years ($n = 411$)	Overall ($n = 3133$)	18–34 years ($n = 1187$)	35–64 years ($n = 1421$)	65 + years ($n = 483$)	Overall sample ($n = 5979$)
Physical activity									
Mild	24.3	24.1	22.1	32.8*	25.6	26.8	23.3	30.1*	24.8
Moderate	27.1	32.2	27.1	14.9*	35.9†	45.7	36.2	10.7*	31.5
Strenuous	13.4	25.3	8.0	1.3*	5.8†	9.9	4.3	0.4*	9.3
None at all	20.9	15.3	20.3	36.1*	18.8	11.2	16.9	43.4*	20.3
Smoking status									
Current regular cigarette smoker	31.2	37.8	29.9	19.6*	30.2	39.5	27.0	16.2*	30.7
Alcohol									
Exceeds limit	27.9	33.1	25.9	16.8*	21.3†	28.7	13.4	18.7*	24.9
Body mass index									
Normal, $<25\ kg\,m^{-2}$	48.5	62.4	38.2	49.7*	67.1†	76.5	61.3	61.8	58.4
Overweight, $25–29.9\ kg\,m^{-2}$	40.0	31.4	47.1	37.4	24.5	18.1	28.7	26.8	31.8
Obese, $>30\ kg\,m^{-2}$	11.5	6.2	14.7	12.9	8.4	5.5	10.0	11.4*	9.9

Mild exercise – ≥4 times per week; moderate exercise – ≥3 times per week; strenuous exercise – ≥ 3 times per week.
* Significant difference between age groups: $P < 0.01$
† Significant difference between males and females: $P < 0.01$.

https://doi.org/10.1079/PHN2004670 Published online by Cambridge University Press

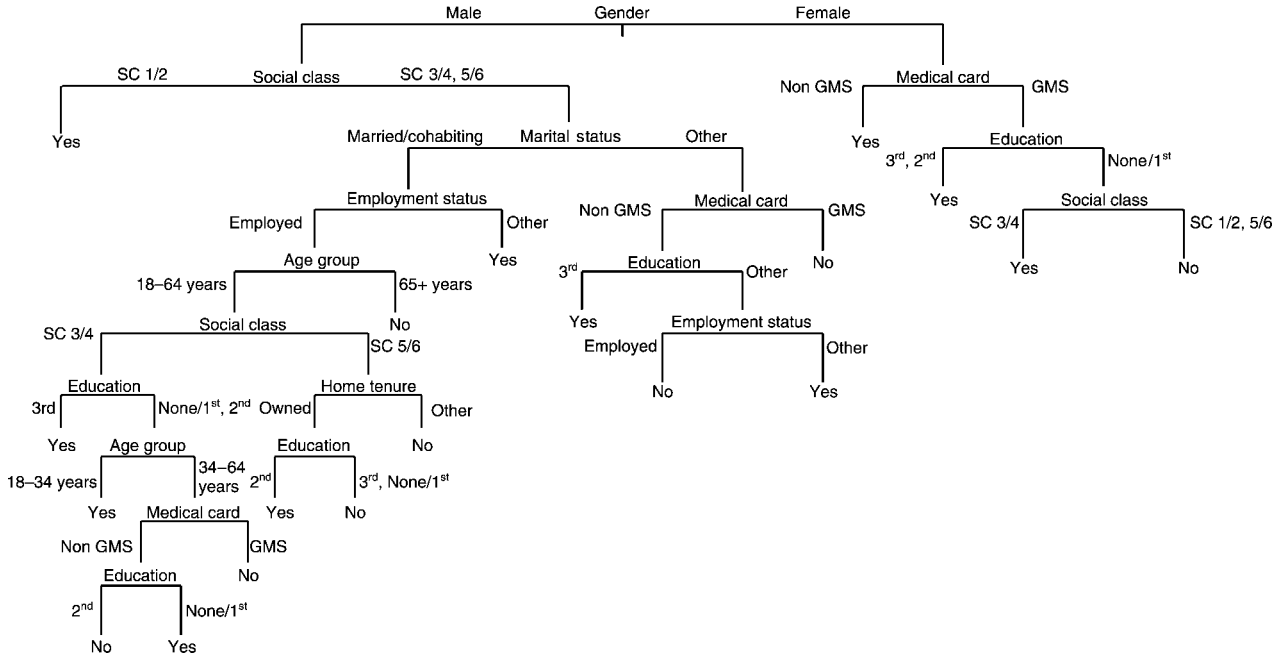

Fig. 1 Pruned classification tree for fruit and vegetable consumption: sociodemographic and socio-economic profile (Yes, ≥4 servings per day; No, <4 servings per day). GMS – General Medical Services Scheme

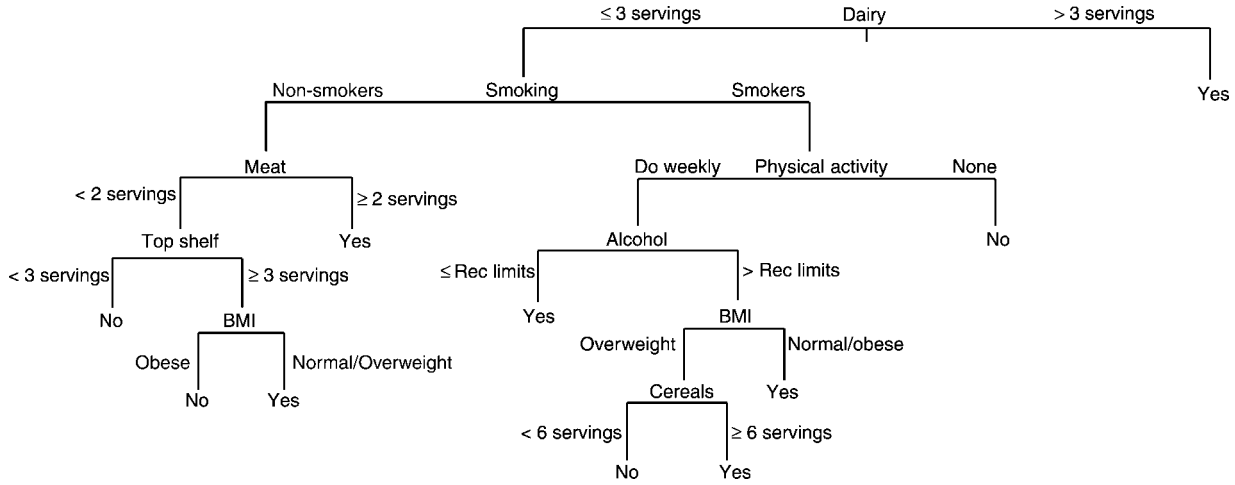

Fig. 2 Pruned classification tree for fruit and vegetable consumption: diet and health-related lifestyle behaviour profile (Yes, ≥4 servings per day; No, < 4 servings per day). BMI – body mass index

recommendation among those who ate less than 3 servings of dairy foods was best predicted by a combination of smoking, alcohol, exercise and other food habits. For those consuming less than the 3 servings of dairy foodstuffs, smoking became an important predictor of compliance with the fruit and vegetable recommendation.

Non-dietary factors were most dominant in the profile of smokers who complied or not with the fruit and vegetable recommendations. Smokers who did no weekly exercise were not likely to eat 4 or more servings of fruit and vegetables daily whereas smokers who did exercise and consumed within the limit of recommended weekly number of alcohol units did. Only among those consuming above the weekly alcohol recommendations did BMI and food in terms of compliance with the cereals, breads and potatoes recommendation become important predictors of fruit and vegetable consumption.

Among non-smokers it was predominantly other dietary behaviours which were important factors in determining compliance with the fruit and vegetable recommendation. Respondents who ate less than 3 servings of dairy foods daily and did not smoke but ate 2 or more servings from the meat, fish and poultry shelf of the pyramid were likely to comply with the fruit and vegetable recommendation. If such respondents ate less than 2 servings of meat, fish and poultry daily and ate 3 or more servings of foods high in fat and/or sugar and were either normal-weight or obese, they too were likely to eat 4 or more servings of fruit and vegetables daily.

https://doi.org/10.1079/PHN2004670 Published online by Cambridge University Press

Discussion

The aim of this paper was to explore in a novel way the social variation in compliance with dietary recommendations of Irish adults. The basic understanding behind any graphical representation of dietary recommendations, such as the food pyramid, is that it is food and not nutrients that people eat, and that by following a suggested balanced combination of foodstuffs to be eaten daily then adequate nutrient availability will generally result[35]. For public health nutrition policy and intervention purposes, groups identified at nutritional risk can be targeted more easily through dietary practice than nutrients[18].

Our results concur with those from a number of studies undertaken in other developed countries[16,18,19,36,37], in that Irish people do not comply with the dietary recommendations but this varies greatly by social circumstance. Bivariate analyses of the SLAN data identified the association between individual social status factors, lifestyle behaviours and dietary choices, particularly consumption of 4 or more servings from the fruit and vegetable shelf of the pyramid. However, the multifaceted nature of nutrition-related behaviour implies that a single explanatory variable is unlikely to capture all dimensions related to correctly classifying individuals in terms of dietary compliance. It was necessary therefore to undertake a more sophisticated multivariate analysis to investigate formally the interactions that exist between sociodemographic, socio-economic and social support-type indicators, and also between other lifestyle behaviours that may affect food choice.

Investigation into the prediction and patterning of dietary habits using classical multivariate techniques such as discriminant analysis and logistic regression has been undertaken in recent years[16,19,36,38]. Whilst the patterning approach may take into account the inter-correlation between dietary components, it does not necessarily account for a large proportion of the total variance in food intake[38]. Thus the introduction of non-dietary factors, such as social and environmental, is necessary to explore more fully the factors influencing dietary intake.

Our analysis used an alternative non-parametric approach, the CART, typically suited to both categorical and continuous-level data with many potential explanatory variables, as in the SLAN dataset. In circumstances such as in SLAN, tree-based approaches have an advantage over the classical methods because they do not have to conform to the same distribution restrictions and are particularly useful when predictors may be associated in some non-linear or non-additive fashion[22–24]. From a statistical perspective, CART is a useful classification and predictive technique and appears well suited to analyses such as those presented here, which require significant interactions to be identified in a manner providing insight and understanding into the structure of the data. The predictive power of the two classification trees is moderately good and highlights the complex nature of the interplay between sociodemographic and socio-economic type variables and food behaviour, specifically in terms of fruit and vegetable consumption prediction. CART analyses of the compliance with fruit and vegetable recommendations appear to be novel within the health and lifestyle behaviour arena.

The classification trees suggest that the most important predictor of fruit and vegetable consumption is gender. Then, as in studies in The Netherlands[16], the USA[17] and Germany[38], quite different social characteristics emerge for males and females in terms of relative importance in predicting consumption of the recommended daily fruit and vegetable servings. For females it was only the socio-economic factors that mattered: medical card eligibility, which is income-related, education and social class. The strong socio-economic status relationship between food choice and females observed in SLAN was also found in a study in Scotland[39] and in older British adults[18], where compliance with dietary recommendations and consumption of a healthy diet were found to be about twice as likely among females, non-manual classes and those from higher-income households.

Among males, a complex constellation of determining factors emerged, with socio-economic, demographic and socially contextual-type factors interweaving with each other at different stages. Blaxter[40] noted that men gain a greater health advantage from being married than women do. The findings by Roos and colleagues[10] support this relationship between family status and food behaviour, with the association being stronger for males than for females. From the tree classification analysis of the SLAN data a similar observation was made for compliance with the recommended daily consumption of fruit and vegetables among males. Clearly it is not just a biochemical sex effect that is being evoked, rather the cultural aspect of gender. Women have traditionally been responsible for controlling the health and related factors of the family. They are more involved in and responsible for the food-related practices of the family[41]. The impact therefore of divorce or becoming a widow may be greater among men in terms of non-adherence to dietary recommendations, a finding observed in SLAN and other similar surveys[10].

As in a study of Norwegian adults[42], other dietary and lifestyle variables, as well as social factors, impact on the consumption of fruit and vegetables among Irish adults. Interestingly, consumption of dairy foodstuffs appears important in determining fruit and vegetable consumption. The idea of supplementation of fat derived from dairy produce with fibre through fruit and vegetable consumption is a very positive behaviour and recommended in many nutrition health promotion messages[39,43]. A combination of non-dietary behaviours showed a consistent pattern of healthier options more likely to lead to compliance with fruit and vegetable recommendations.

https://doi.org/10.1079/PHN2004670 Published online by Cambridge University Press

There did, however, appear to be a compensatory element between the variables, particularly around smoking, with for example smokers who did regular exercise consuming the recommended servings of fruit and vegetables. This suggestion that a combination of healthy and unhealthy lifestyle behaviours coexists agrees with work in The Netherlands, which found no exclusive lifestyle for health risk[16].

Study limitations

The tendency of long food-frequency questionnaires (FFQs) to overestimate[31] and the effect of potential sampling and response biases have to be taken into consideration. The potential for non-response bias and selection bias is always a concern in self-administered postal surveys. However, a very good response rate for a postal questionnaire of its type was achieved, suggesting confidence in the survey's representation of the Irish adult population. There was good completion of most questions, although classification into social class was relatively low due to a third of respondents not providing occupational details.

FFQs have been used regularly to determine population dietary habits[44–46]. While there are well-known limitations in the use of an FFQ for assessing individual nutrient intake, as a tool for population ranking and discrimination between groups in large-scale dietary surveys, the FFQ is the most feasible and useful[47–49]. A thorough validation of the SQFFQ used in the British arm of EPIC has been undertaken in several populations[27], and likewise the adapted Irish version was validated using food diaries and urinary protein with *p*-aminobenzoic acid in staff and students of the National University of Ireland, Galway[50].

Social desirability, differing ability to estimate frequency of consumption of foodstuffs and literacy levels may all contribute to misclassification of social groups. Social desirability bias, with overreporting of what is socially desirable, underreporting of what is not and the possible confounding of the nature of relationships between the variables under study, can arise through the reliance on self-reported behaviour[51]. It is acknowledged that such a bias is possibly prevalent in the SLAN dataset. However, the social patterns observed using this approach concur strongly with those reported using other dietary assessment methods, suggesting that the prevailing social inequalities in compliance with dietary recommendations are greater than the error associated with the assessment method.

Conclusions

This novel application of a non-hierarchical statistical approach has clearly demonstrated the existence of social inequalities in current Irish dietary patterns. Based on these empirical findings and the solid methodological rationale presented for its application, the tree classification system is a useful method that should be more widely applied.

Very marked gender differences were observed in dietary behaviour and in the relative importance of different social status factors on compliance with the daily recommendation for fruit and vegetable consumption. Material and structural influences, reflected through social class, education and medical card status (an indicator of financial means), matter very much for females. For males, while these factors are important, they appear to be mediated through other more socially contextual-type factors such as being married and the related household socio-economic circumstances. Clearly there are a number of interrelated factors, structural, material and psychosocial in nature, that together contribute to the dietary choices adults make. Recognition of the role that each of these compositional and contextual factors plays in influencing dietary habits of men and women has implications for the manner in which dietary strategies and policies are developed and implemented. To impact on the population's health, action is necessary at the macro, meso and micro level. A concerted health promotion approach encompassing the various possible determining factors is one which aims to reduce macrosocial inequality through adequate policy initiatives and infrastructures that facilitate affordable food choices, combined with informed, targeted education in various settings.

Acknowledgements

The 1998 Survey of Lifestyle, Attitudes and Nutrition was commissioned and funded by the Health Promotion Unit of the Department of Health and Children, Republic of Ireland. The design and data collection phase was overseen by a multi-discipline steering committee comprised of academic members, health board representatives and members of the Department of Health and Children.

References

1 Beer-Borst S, Hercberg S, Morabia A, Bernstein MS, Galan P, Galasso R, *et al*. Dietary patterns in six European populations: results from EURALIM, a collaborative European data harmonisation and information campaign. *European Journal of Clinical Nutrition* 2000; **54**: 253–62.
2 James WPT, Nelson M, Ralph A, Leather S. The contribution of nutrition to inequalities in health. *British Medical Journal* 1997; **314**: 1545–9.
3 Leather S. Fruit and vegetables: consumption patterns and health consequences. *British Food Journal* 1995; **97**: 10–17.
4 McElduff P, Dobson AJ. Trends in coronary heart disease – has the socioeconomic differential changed? *Australian and New Zealand Journal of Public Health* 2000; **24**: 465–73.
5 Davey Smith G, Marmot M. Trends in mortality in Britain: 1920–1986. *Annals of Nutrition & Metabolism* 1991; **35**(Suppl. 1): 53–63.
6 Milligan RAK, Burke V, Beilin LJ, Dunbar DL, Spencer MJ, Balde E, *et al*. Influence of gender and socioeconomic status on dietary patterns and nutrient intakes in 18 year old

https://doi.org/10.1079/PHN2004670 Published online by Cambridge University Press

Australians. *Australian and New Zealand Journal of Public Health* 1998; **22**: 485–93.
7 Billson H, Pryer JA, Nichols R. Variation in fruit and vegetable consumption among adults in Britain. An analysis from the dietary and nutritional survey of British adults. *European Journal of Clinical Nutrition* 1999; **53**: 946–52.
8 Popkin BM, Siega-Riz AM, Haines PS. A comparison of dietary trends among racial and socioeconomic groups in the United States. *New England Journal of Medicine* 1996; **335**: 716–20.
9 Mishra G, Ball K, Arbuckle J, Crawford D. Dietary patterns of Australian adults and their association with socioeconomic status: results from the 1995 National Nutrition Survey. *European Journal of Clinical Nutrition* 2002; **56**: 687–93.
10 Roos E, Lahelma E, Virtanen M, Prattala R, Pietinen P. Gender, socioeconomic status and family status as determinants of food behaviour. *Social Science & Medicine* 1998; **46**: 1519–29.
11 Beaton GH. Approaches to analysis of dietary data: relationship between planned analyses and choice of methodology. *American Journal of Clinical Nutrition* 1994; **59**: 253S–61S.
12 Fidanza F, Gentile MA, Porrini M. A self administered semiquantitative food frequency questionnaire with optical reading and its concurrent validation. *European Journal of Epidemiology* 1995; **11**: 163–70.
13 Nutrition Advisory Group to the Department of Health (NAG). *Recommendations for a Food and Nutrition Policy*. Dublin: NAG, 1995.
14 Willett WC, Sacks F, Trichopoulou A, Drescher G, Ferro-Luzzi A, Helsing E, *et al*. Mediterranean diet pyramid: a cultural model for healthy eating. *American Journal of Clinical Nutrition* 1995; **61**: 1402S–6S.
15 Hermann-Kunz E, Thamm M. Dietary recommendations and prevailing food and nutrient intakes in Germany. *British Journal of Nutrition* 1999; **81**: S61–9.
16 Hulshof KFAM, Wedel M, Lowik MRH, Kok FJ, Kistemaker C, Hermus RJJ, *et al*. Clustering of dietary variables and other lifestyle factors (Dutch Nutrition Surveillance System). *Journal of Epidemiology and Community Health* 1992; **46**: 417–24.
17 Wirfalt AKE, Jeffery RW. Using cluster analysis to examine dietary patterns: nutrient intakes, gender and weight status differ across food pattern clusters. *Journal of American Dietetic Association* 1997; **97**: 272–9.
18 Pryer JA, Nichols R, Elliott P, Thakrar B, Brunner E, Marmot M. Dietary patterns among a national random sample of British adults. *Journal of Epidemiology and Community Health* 2001; **51**: 29–37.
19 Martikainen P, Brunner E, Marmot M. Socioeconomic differences in dietary patterns among middle aged men and women. *Social Science & Medicine* 2003; **56**: 1397–410.
20 McCann SE, Weiner J, Graham S, Freudenheim JL. Is principal components analysis necessary to characterise dietary behaviour in studies of diet and disease? *Public Health Nutrition* 2001; **4**: 903–8.
21 Ursin G, Ziegler RG, Subar A, Graubard BI, Haile RW, Hoover R. Dietary patterns associated with a low fat diet in the National Health Examination Follow up Study: identification of potential confounders for epidemiological analyses. *American Journal of Epidemiology* 1993; **137**: 916–27.
22 Breiman LJH, Friedman R, Olshen A, Stone CJ. *Classification and Regression Trees*. Monterey, CA: Wadsworth Inc., 1984.
23 Steinberg D, Colla P. *CART: Tree-structured Non Parametric Data Analysis*. San Diego, CA: Salford Systems, 1995.
24 Bloch DA, Segal MR. Empirical comparison of approaches of forming strata: using classification trees to adjust for covariates. *Journal of the American Statistical Association* 1989; **84**: 897–905.
25 Friel S, Kelleher CC, Nolan G, Harrington J. Social diversity of Irish adults nutritional intake. *European Journal of Clinical Nutrition* 2003; **57**: 865–75.
26 Riboli E, Kaaks R. The EPIC project: rationale and study design. *International Journal of Epidemiology* 1997; **26**(Suppl. 1): S6–14.
27 Bingham SA, Gill C, Welch A, Cassidy A, Runswick SA, Oakes S, *et al*. Validation of the dietary assessment methods in the UK arm of EPIC using 24-hour urinary nitrogen and potassium and serum vitamin C and carotenoids as biomarkers. *International Journal of Epidemiology* 1997; **26**(Suppl. 1): S137–51.
28 Centre for Health Promotion Studies (CHPS). *The National Health and Lifestyle Surveys (SLAN/HBSC)*. Galway: CHPS, National University of Ireland, 1999.
29 Central Statistics Office. *Census of the Population 1996*. Dublin: The Stationery Office, 1996.
30 Willett WC. Issues in analysis and presentation of dietary data. In: Willett WC, ed. *Nutritional Epidemiology*, 2nd ed. Oxford: Oxford University Press, 1998; 398–9.
31 Burley V, Cade J, Margetts B, Thompson R, Warm D. *Consensus Document on the Development, Validation and Utilisation of Food Frequency Questionnaires*. Leeds/Southampton: Nuffield Institute for Health, University of Leeds & Institute of Human Nutrition, University of Southampton, 2000.
32 Clarke GM, Cooke D. *A Basic Course in Statistics*, 4th ed. London: Arnold, 1998.
33 Department of Health (DoH). *Nutrition Health Promotion Framework for Action*. Dublin: DoH, 1991.
34 Hjorth JSU. *Computer Intensive Statistical Methods Validation, Model Selection, and Bootstrap*. London: Chapman & Hall, 1994.
35 Yates AA. Process and development of dietary reference intakes: basis, need and application of recommended dietary allowances. *Nutrition Reviews* 1998; **56**: S5–9.
36 Pryer JA, Cook A, Shetty P. Identification of groups who report similar patterns of diet among a representative national sample of British adults aged 65 years of age or more. *Public Health Nutrition* 2001; **4**: 787–95.
37 Johansson L, Andersen LF. Who eats 5 a day?: intake of fruit and vegetables among Norwegians in relation to gender and lifestyle. *Journal of the American Dietetic Association* 1998; **98**: 689–91.
38 Schulze MB, Hoffmann K, Kroke A, Boeing H. Dietary patterns and their association with food and nutrient intake in the European Prospective Investigation into Cancer and Nutrition (EPIC)–Potsdam study. *British Journal of Nutrition* 2001; **85**: 363–73.
39 Anderson A, Hunt K. Who are the healthy eaters? Eating patterns and health promotion in the west of Scotland. *Health Education Journal* 1992; **51**: 3–10.
40 Blaxter M. *Health and Lifestyles*. London: Tavistock/Routledge, 1990.
41 Schafer RB, Schafer E. Relationship between gender and food roles in the family. *Journal of Nutrition Education* 1989; **21**: 119–26.
42 Johansson L, Thelle DS, Solvoll K, Bjornboe GEA, Drevon CA. Healthy dietary habits in relation to social determinants and lifestyle factors. *British Journal of Nutrition* 1999; **81**: 211–20.
43 Depatment of Health and Children. *The National Health Promotion Strategy 2000–2005*. Dublin: Stationery Office, 2000.
44 Block G, Subar AF. Estimates of nutrient intake from a food frequency questionnaire: the 1987 national health interview survey. *Journal of the American Dietetic Association* 1992; **92**: 969–77.

45 Johansson L, Solvoll K, Bjorneboe GEA, Drevon CA. Dietary habits among Norwegian men and women. *Scandinavian Journal of Nutrition* 1997; **41**: 63–70.
46 Patterson RE, Kristal AR, Tinker LF, Carter RA, Bolton MP, Agurs-Collins T. Measurement characteristics of the Women's Health Initiative food frequency questionnaire. *Annals of Epidemiology* 1999; **9**: 178–87.
47 Hupkens CLH, Knibbe RA, Drop MJ. Social class differences in women's fat and fibre consumption: a cross national study. *Appetite* 1997; **28**: 131–49.
48 Rimm EB, Giovannucci EL, Stampfer MJ, Colditz GA, Litin LB, Willett W. Reproducibility and validity of an expanded self-administered semiquantitative food frequency questionnaire among male health professionals. *American Journal of Epidemiology* 1992; **135**: 1114–26.
49 Willett WC. Future directions in the development of food frequency questionnaires. *American Journal of Clinical Nutrition* 1994; **59**: 171S–4S.
50 Harrington J. *Validation of food frequency questionnaire for national health and lifestyle survey*. Minor thesis as part MA in Health Promotion, National University of Ireland, Galway, 1998.
51 Phillips DL, Clancy KJ. Some effects of social desirability in survey studies. *American Journal of Sociology* 1972; **77**: 921–38.