Determinants of fruit and vegetable intake in England: a re-examination based on quantile regression

Georgios Boukouvalas, Bhavani Shankar* and W Bruce Traill

Department of Agricultural & Food Economics, University of Reading, PO Box 237, Reading RG6 6AR, UK

Abstract

Objective: To examine the sociodemographic determinants of fruit and vegetable (F&V) consumption in England and determine the differential effects of socio-economic variables at various parts of the intake distribution, with a special focus on severely inadequate intakes.

Design: Quantile regression, expressing F&V intake as a function of socio-demographic variables, is employed. Here, quantile regression flexibly allows variables such as ethnicity to exert effects on F&V intake that vary depending on existing levels of intake.

Setting: The 2003 Health Survey of England.

Subjects: Data were from 11 044 adult individuals.

Results: The influence of particular sociodemographic variables is found to vary significantly across the intake distribution. We conclude that women consume more F&V than men; Asians and blacks more than whites; co-habiting individuals more than single-living ones. Increased incomes and education also boost intake. However, the key general finding of the present study is that the influence of most variables is relatively weak in the area of greatest concern, i.e. among those with the most inadequate intakes in any reference group.

Conclusions: Our findings emphasise the importance of allowing the effects of socio-economic drivers to vary across the intake distribution. The main finding, that variables which exert significant influence on F&V intake at other parts of the conditional distribution have a relatively weak influence at the lower tail, is cause for concern. It implies that in any defined group, those consuming the least F&V are hard to influence using campaigns or policy levers.

Increased consumption of fruit and vegetables (F&V) is at the heart of healthy eating campaigns around the world. The WHO's Global Strategy on Diet, Physical Activity and Health, endorsed by member countries in 2004, includes a global quantitative norm that per capita F&V consumption should exceed 400 g/d. In the UK, this aggregate norm of 400 g/d has been broken down into '5-a-day' portions of at least 80 g each by the Department of Health. The government has invested substantially in promoting the 5-a-day programme, elements of which include the National School Fruit Scheme and a communication strategy incorporating the 5-a-day logo.

In keeping with this interest in F&V consumption, a large volume of research has been produced analysing the determinants of F&V intakes by individuals or households. These range from studies involving analysis of large samples drawn from representative populations, to purposively collected information on population sub-groups of interest. Research, depending on the available data, has alternatively attempted to relate F&V intakes to sociodemographic, socio-economic, psychological or sensory variables. Often, studies have restricted themselves to statistical comparisons of F&V consumption across demographic or socio-economic groups, in an attempt to identify sub-populations for policy targeting. Sometimes, regression-based approaches have been undertaken to identify sources of independent variation in F&V consumption and to quantify the magnitude of the effect exerted by a particular causal variable. Seeking sources of independent variation is important for a more nuanced understanding of the drivers of intake. For instance, a comparison of intake differences across groups in some research[1,2] shows that F&V intake is lower in Scotland and the North of England compared with the rest of the UK. A regression-based examination can reveal whether this effect remains after controlling for differences in incomes, family sizes, etc. across these regions.

In the present research, we examine such independent sources of variation in F&V consumption in England using sociodemographic data from the 2003 Health Survey of England (HSE). In doing so, we employ a powerful technique in the form of quantile regression. This enables

Corresponding author: Email b.shankar@reading.ac.uk

us to explore important but hitherto understudied questions relating to whether and how variables such as income, ethnicity and regional location differentially influence various parts of the conditional distribution of F&V intake, i.e. how the effect on those with severely inadequate intake differs from those with average intakes and those with high intakes.

Regression approaches to determinants of fruit and vegetable intake

Previous regression-based approaches in the literature have mostly employed either multiple linear[3,4] or logistic regressions[1,5,6]. Multiple linear regressions express the conditional mean of F&V intake as a function of independent variables. However, they treat various parts of the conditional distribution of F&V consumption identically and constrain the marginal effects of independent variables to be the same throughout the distribution of F&V consumption, which is a significant shortcoming in nutritional intake problem-setting. From a public health and nutrition policy perspective, particular parts of nutrient intake distributions are likely to be of more interest than others. In the F&V case, the behaviour of those in the lower tail of the conditional intake distribution, e.g. those in the bottom 10 %, is likely to be of more interest than those in the higher reaches of the distribution. It is natural to be more concerned about the 'worst performers' in any performance-related grouping. Most importantly, the effect of an independent variable may reasonably be hypothesised to vary across the distribution of the dependent variable. For instance, given a specification of other conditioning variables, it is likely that the effect of increased education on F&V intake for those currently in the bottom 10 % of the intake distribution will be different from those in the middle, or in the top 10 %.

Limited dependent variable approaches, such as logistic regressions, express probabilities of exceeding cut-off points as a function of covariates. By doing so, they offer a potential way to train focus on particular segments of the intake distribution. For example, the cut-off may be 5-a-day and a logistic regression would model the influence of a socio-economic variable on the probability of exceeding this cut-off (previous studies have typically opted for such a simple binary classification). Or, one may divide the sample up into 'low' (e.g. less than 2-a-day), 'medium' (between 2- and 4-a-day) and 'high' (more than 4-a-day), and ordered logistic regression would estimate the effect of a socio-economic variable on the probability of belonging to each of these groups. Note, however, that this common practice of reducing continuous information on intakes into information on a small number of categories involves a statistical loss of information. For instance, logistic regression involving a category such as 'less than 5-a-day' would treat a data point with intake of 4·9 identically to a data point with intake of 0. As Marmot[7] notes in the context of disease modelling: 'Clinicians tend to view disease in terms of binary opposition… the right question [is] not whether a person has the disorder or not, but how much of it does he have. More often than not, both the exposures (causes) and the outcomes (effects) are distributed continuously'. Note also that, although logistic regression can help obtain focus on particular segments of the intake distribution, the effect of a socio-economic variable on intake remains constant regardless of what the existing level of intake is.

Quantile regression[8] (QR) in this context would allow the impact of the explanatory variable to vary along the whole range of F&V intake ('quantile' is general terminology for what may be referred to as percentile, decile, quartile, etc. in specific cases). QR methods have gained popularity among economists and ecologists over the last decade. They hold particular promise in applications to nutrition problems where dietary excess and/or inadequacy questions beg particular attention to the tails of distributions, although there seem to be only a small number of applications so far[9,10]. Accessible introductions to QR methods are available in Koenker and Hallock[11] and Cade and Noon[12]. Some additional statistical explanation is presented in the Appendix.

Methods

Data and variables

The present study uses data from the HSE undertaken in 2003[13]. The HSE 2003 is the thirteenth annual survey of its kind conducted by the Department of Health, UK, with the objective of monitoring the nation's health. Each survey in the series consists of a series of modules wherein some questions are common each year, while others are repeated at regular intervals. The HSE 2003 survey included questions about F&V consumption for informants aged 5 years and over. The survey was designed to provide a representative sample of the population of all ages in private households in England. Interviews were obtained with 14 836 adults (aged 16 years and over) and 3717 children (aged under 16 years), resident in 8867 households. The estimated response rate was 66 %.

Surveys were done in-person, involving personal interviews, physical measurements and nurse visits. A multistage stratified sampling design was adopted for the selection of households, using postcode address files as the sampling frame, and stratifying on the basis of local authority and the percentage of households with non-manual *v.* manual household head occupations within the postcode. In most cases, data were collected on all members of a household, although in households with three or more children, two children were randomly selected for interview in order to reduce household interview burden. F&V intake information was collected

Table 1 Summary statistics for continuous variables in the sample: data from 11 044 adult individuals in the 2003 Health Survey of England (HSE)

Variable name	Variable description	Mean	Median	SD
F&V	F&V portions	3·5	3·0	2·5
CHILDREN	Number of children	0·6	0	1·0
BMI	BMI (kg/m^2)	27·0	26·4	5·0
AGE	Age (years)	47·5	46·0	18·0
EQVINCOME	Equivalised income* (1000s of £)	26·6	19·5	23·8

*'Equivalised income' is the HSE 2003 income measure used here. Ordinary household income data are reported in the data set only as a categorical variable with broad bands. HSE 2003 calculates a McClement score for each household (a measure that depends on number, age and relationships of adults and children in the household) and divides the raw income data by the McClement score to arrive at equivalised income. More details are available elsewhere[14].

based on recall over the 24 h ending the previous mid-night of the interview. Details of the data collection methodology are available elsewhere[14].

For the present study, after deleting observations with missing values, data on 11 044 adults were retained for analysis. F&V consumption in (80 g) portions per day was designated the dependent variable, while the independent variable set comprised a variety of sociodemographic variables that have been associated with F&V intake in previous literature. The 80 g portions per day is used here instead of direct expression in grams because the '5-a-day' programme has resulted in number of portions per day coming into widespread use. In some cases, the original data were transformed to create new variables for ease of interpretation and comparison with previous studies. Table 1 shows explanations and summary statistics of the continuous variables used in the analysis.

Several independent variables are of a categorical nat-ure, and for convenience in presentation, some of the category sets were redefined and reduced to a smaller number of categories. The defined categories and cate-gory-wise break up of the sample was as follows. Gender: female, 46%; male, 54%. Co-habitation: co-habiting (living with a partner/spouse), 56%; single, 44%. Highest educational attainment: no qualifications, 23%; up to GCSE, 32%; GCSE to A-levels (more than GCSE, up to and including A-levels), 29%; A-levels to degree or above (more than A-levels, including degree or more), 16%. Social class:* routine and manual, 40%; intermediate, 19%; managerial or professional, 41%. Race: white, 94%; black, 2%; Asian, 3%; other/mixed, 1%. Region: York-shire & Humber, 9%; North, 21%; West Midlands, 11%; East, 49%; South West, 10%. Location: suburban, 58%; urban, 17%; rural, 25%. Season during which F&V con-sumption recorded: autumn, 27%; spring, 25%; summer, 26%; winter, 22%. Self-reported health status: good, 76%;

fair or bad, 24%. For these categorical independent variables, a number of dummy variables were defined for use in regression analysis. In each of the variables in the category list above, the first category was used as the 'base' in defining dummies to measure changes against. For instance, 'Female' was the base for gender and coded as 0, and so the gender dummy variable is called 'Male' with men coded as 1.

Regression analyses

A multiple linear regression was first estimated using Ordinary Least Squares (OLS) to provide a basis for comparison with the quantile regression. The functional form for all regressions was initially specified as quadratic in all continuous independent variables (i.e. income, household size, total children, BMI and age). This would allow the effect of e.g. household size on F&V intake to grow or dampen as household size rises. The quadratic terms for both the total children as well as the BMI vari-ables were consistently found to be insignificantly dif-ferent from zero in all regressions, and were subsequently dropped. The continuous variables were centred at their medians to assist interpretation.

In the QR, seventy-three different conditional quantile functions were estimated for F&V intake, starting with the 0·05 (5%) quantile and proceeding in 0·0125 increments until the 0·95 (95%) quantile (i.e. 0·05, 0·0625, 0·75, 0·875, 0·1, …, 0·95). The starting and finishing points were set at 0·05 and 0·95 because extreme quantiles are known to encounter problems with stable confidence interval estimation. A Markov chain marginal bootstrap[15] was implemented to compute confidence intervals for every quantile estimate. All programming was imple-mented in the SAS© statistical software package version 9·1 (SAS Institute, Inc., Cary, NC, USA).

Results

As shown in Table 1, the mean F&V intake in the sample is 3·5 portions/d, well short of the 5-a-day mark. The standard deviation of 2·5 indicates large variance in intakes, which is confirmed by the histogram in Fig. 1. More than 10% of the sample is extremely deficient in

* The HSE uses the UK National Statistics Socio-Economic Classification. The broadest version is the three-class version used here. The 'routine and manual' class includes long-term unemployed, routine occupations (those involving basic labour contracts, with little need for employee discretion), semi-routine occupations (where employers offer slightly better than basic labour contracts, with some employee discretion) and lower supervisory and technical occupations (e.g. foremen, supervisors, those with some level of work autonomy). Further details are available from the Office of National Statistics, UK (www.statistics.gov.uk).
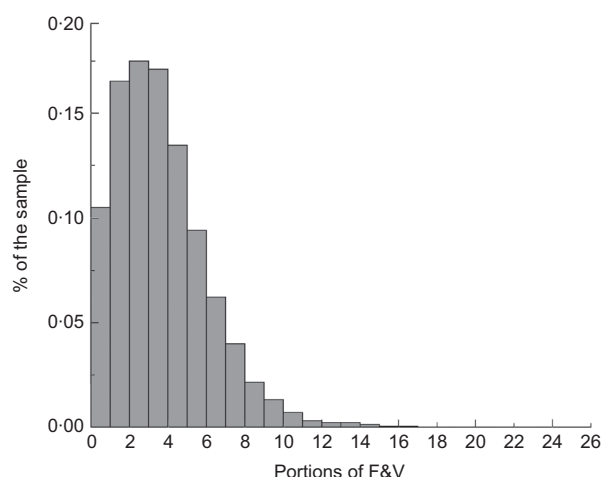
Fig. 1 Distribution of fruit and vegetable (F&V) consumption (in 80 g portions/d) in the sample: data from 11 044 adult individuals in the 2003 Health Survey of England

F&V consuming less than 1 portion/d. Fully half the sample has an intake of 3 portions/d or less. This large variance in intakes involving a substantial likelihood of deficiency suggests that the standard regression proposition, that F&V intake response to a socio-economic driver is the same regardless of the baseline intake level, is unlikely to be a realistic representation.

With twenty-eight independent variables and estimates available for seventy-three different conditional quantiles, presentation of results necessarily has to be selective. Graphs plotting estimates against quantiles for individual variables can adequately capture and summarise the cross-quantile variation in estimates. We employ such graphs for a selection of eight independent variables, selected either because of their importance in explaining F&V intake in previous literature or because they demonstrate interesting variation across different levels of F&V intake. These graphs (Figs 2 to 9 below) show the quantile estimates along with the bootstrapped 95 % confidence intervals shown as shaded areas around the estimates. In addition, the OLS estimate and the 95 % confidence interval are also shown as dashed and dotted straight lines, respectively, superimposed on the QR plots. Table 2 presents tabulated results for all variables, for a selected set of quantiles: 0·05 (the lowest conditional quantile function estimated), 0·25, 0·50 (median), 0·75 and 0·95 (the highest conditional quantile function estimated). The final columns of Table 2 also show the multiple linear regression (OLS) estimates.

It is readily apparent from Figs 2 to 9 that in most cases at least some portion of the QR estimates lie outside the OLS confidence intervals. This suggests that the simple conditional mean shift implied by the OLS model is not plausible. The first socio-economic driver we discuss is gender, which has been linked to F&V intake in previous studies. In the UK, the indication has been that not only
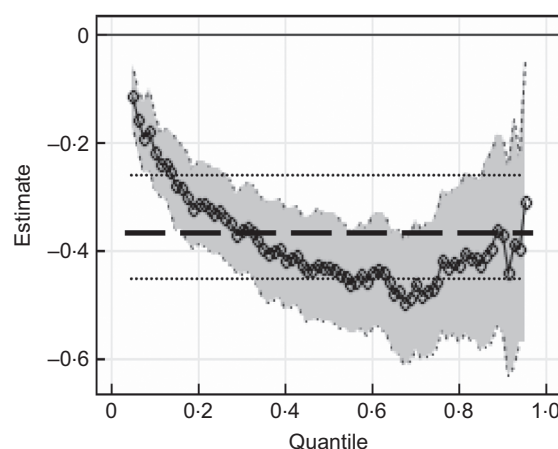

Fig. 2 Effect of gender (men in comparison to women) on fruit and vegetable (F&V) consumption (in 80 g portions/d): data from 11 044 adult individuals in the 2003 Health Survey of England. Quantile regression estimates across quantiles of the F&V intake distribution (–○–) with the bootstrapped 95 % confidence intervals shown as shaded areas around the estimates; Ordinary Least Squares estimates (— — —) and the corresponding 95 % confidence intervals (·······) are also shown
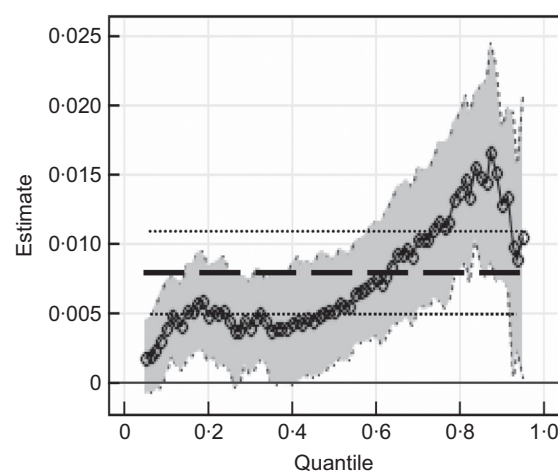

Fig. 3 Effect of equivalised income (1000s of £) on fruit and vegetable (F&V) consumption (in 80 g portions/d): data from 11 044 adult individuals in the 2003 Health Survey of England. Quantile regression estimates across quantiles of the F&V intake distribution (–○–) with the bootstrapped 95 % confidence intervals shown as shaded areas around the estimates; Ordinary Least Squares estimates (— — —) and the corresponding 95 % confidence intervals (·······) are also shown

do women eat more F&V[1], but they are also attitudinally better disposed to incorporating more F&V in their diets[16]. Our results confirm this, as shown in Fig. 2. The OLS results show that, after controlling for changes in other variables, males consume approximately 0·37 portions (27 g) per day less than women. This is consistent with the 0·3 portion difference reported by Doyle and Hosfield[1] based on the HSE 2001, although their analysis did not control for other variables. However, consideration of the QR results presented in Fig. 2 shows that there
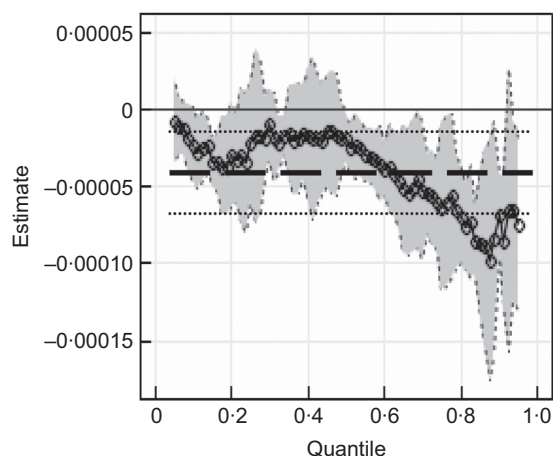
Fig. 4 Effect of equivalised income (1000s of £) squared on fruit and vegetable (F&V) consumption (in 80 g portions/d): data from 11 044 adult individuals in the 2003 Health Survey of England. Quantile regression estimates across quantiles of the F&V intake distribution (–○–) with the bootstrapped 95 % confidence intervals shown as shaded areas around the estimates; Ordinary Least Squares estimates (— — —) and the corresponding 95 % confidence intervals (·······) are also shown

Fig. 6 Effect of educational attainment (GCSE to A-levels compared with no qualifications) on fruit and vegetable (F&V) consumption (in 80 g portions/d): data from 11 044 adult individuals in the 2003 Health Survey of England. Quantile regression estimates across quantiles of the F&V intake distribution (–○–) with the bootstrapped 95 % confidence intervals shown as shaded areas around the estimates; Ordinary Least Squares estimates (— — —) and the corresponding 95 % confidence intervals (·······) are also shown
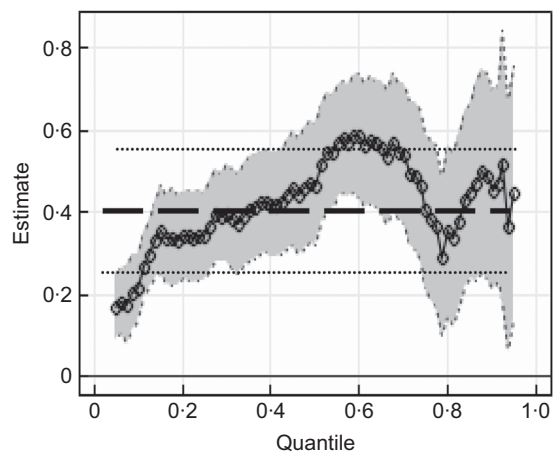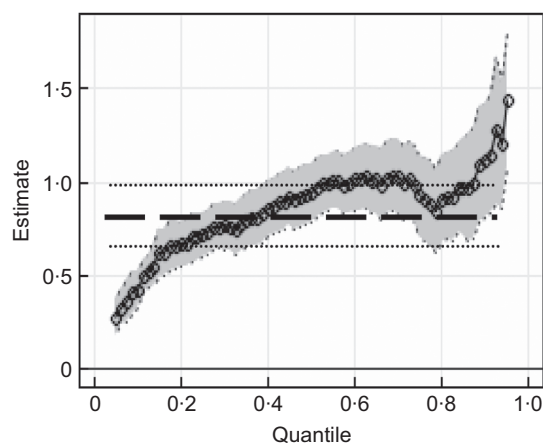

Fig. 5 Effect of educational attainment (up to GCSE compared with no qualifications) on fruit and vegetable (F&V) consumption (in 80 g portions/d): data from 11 044 adult individuals in the 2003 Health Survey of England. Quantile regression estimates across quantiles of the F&V intake distribution (–○–) with the bootstrapped 95 % confidence intervals shown as shaded areas around the estimates; Ordinary Least Squares estimates (— — —) and the corresponding 95 % confidence intervals (·······) are also shown
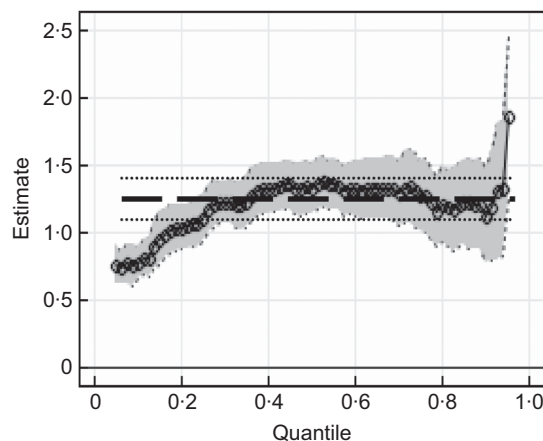

Fig. 7 Effect of educational attainment (A-levels to degree compared with no qualifications) on fruit and vegetable (F&V) consumption (in 80 g portions/d): data from 11 044 adult individuals in the 2003 Health Survey of England. Quantile regression estimates across quantiles of the F&V intake distribution (–○–) with the bootstrapped 95 % confidence intervals shown as shaded areas around the estimates; Ordinary Least Squares estimates (— — —) and the corresponding 95 % confidence intervals (·······) are also shown

is significant variation in the gender effect across the intake distribution. At low levels of F&V intake, the gender difference is substantially less marked, with the difference being only 0·11 portions (about 9 g) at the 0·05 quantile. The gender effect peaks at almost half a portion (40 g) around the 0·7 percentile, gently declining in the higher intake area.

The OLS regression as well as QR results in Figs 3 and 4 at any given quantile suggest that increasing income

generally increases F&V intake, but that this effect levels off as income gets larger. This is consistent with evidence from previous literature[17,18]. Since the continuous variables have been centred to the median, the parameter estimates of the level (non-squared) terms indicate the marginal effect at the sample median. Thus at the median income in the sample, the OLS results indicate that every £1000 increase in income results in increased F&V intake of 0·007 portions, or approximately 0·6 g. Thus the
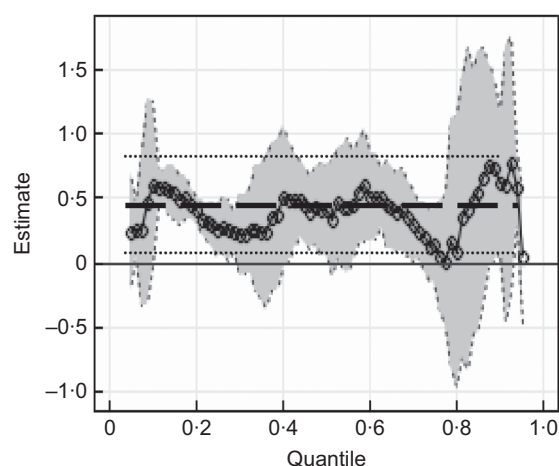
Fig. 8 Effect of ethnicity (black compared with white) on fruit and vegetable (F&V) consumption (in 80 g portions/d): data from 11 044 adult individuals in the 2003 Health Survey of England. Quantile regression estimates across quantiles of the F&V intake distribution (–○–) with the bootstrapped 95 % confidence intervals shown as shaded areas around the estimates; Ordinary Least Squares estimates (— — —) and the corresponding 95 % confidence intervals (·······) are also shown
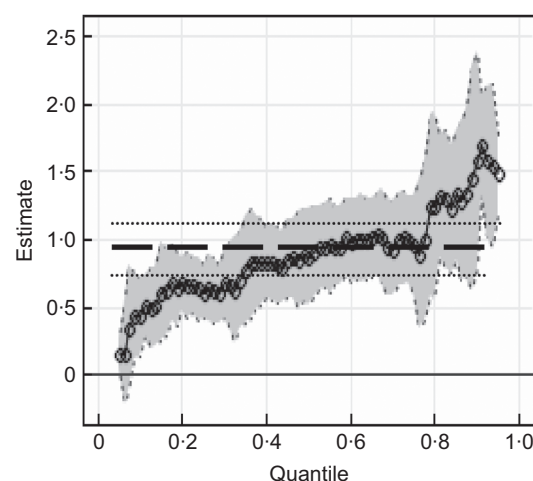

Fig. 9 Effect of ethnicity (Asian compared with white) on fruit and vegetable (F&V) consumption (in 80 g portions/d): data from 11 044 adult individuals in the 2003 Health Survey of England. Quantile regression estimates across quantiles of the F&V intake distribution (–○–) with the bootstrapped 95 % confidence intervals shown as shaded areas around the estimates; Ordinary Least Squares estimates (— — —) and the corresponding 95 % confidence intervals (·······) are also shown

independent income effect appears to be small, although statistically significant. However, QR shows that this income effect can be even smaller at the lower end of the F&V consumption distribution. At the extreme, for the 0·05 quantile, the computed effect of every £1000 increase from the median income level is only 0·0017 portions (0·13 g) of increased F&V intake. Indeed, the effect at this quantile is statistically insignificantly different from zero.

There exists a significant body of evidence indicating that education has a strong influence on F&V consumption. For example, Ricciuto *et al.*[17] used multiple regression on data from Canada to show that a household where the reference person has a university degree purchases 14 % more F&V than one with only basic schooling. Figures 5 to 7 clearly demonstrate the strong influence of education on F&V intake in England. The OLS results indicate that, holding all else constant, an individual with GCSE qualifications consumes about 0·4 portions (32 g) per day more than the base case person with no qualifications. A-level and degree qualifications raise this to approximately 0·8 (64 g) and 1·2 (96 g) portions respectively, compared with the base case. These are significant effects, but the QR results shown in the figures indicate that the OLS results tell only part of the story. The effects are substantially dependent on the part of the F&V intake distribution at which they are computed. Generally, very strong effects at higher levels of intake counterbalance relatively weak effects at lower levels. For example, A-level educated individuals at the 0·95 quantile consume almost 1·5 portions (120 g) more than those with no qualifications all else equal, but this education effect

drops to only 0·28 portions (8·5 g) at the 0·05 quantile. In each of the higher educational attainment categories compared with the base case of no qualifications, the effect of the higher education on the bottom fifth of the conditional F&V intake distribution is comparatively weak.

Successive food surveys conducted in the UK have reported that black and Asian minority groups consume more F&V than the general population. The British Heart Foundation[19] notes that vegetable consumption is highest in the Asian and Chinese sub-populations, while fruit consumption is highest in the black and mixed categories. The OLS regression results reported here indicate that, after controlling for other sources of variation, blacks on average consume 0·4 portions (32 g) of F&V more than the base white case, while Asians consume almost a full portion more (0·9 portions, 72 g). The QR results in Figs 8 and 9 reveal that while the effect observed among blacks is relatively constant across quantiles (and constitutes a rare case where the QR results are more or less entirely within OLS confidence intervals), the effect observed among Asians is very quantile-dependent. Thus, at very high intake levels (0·95 quantile), the effect observed among Asians amounts to almost 1·5 portions more compared with whites, while at severe intake deficiency levels (0·05 quantile) this effect amounts to only 0·15 portions. In other words, there is a tenfold difference in the effect observed among Asians across the F&V intake distribution.

An important pattern appears to emerge from Figs 2 to 9 and the associated discussion above. Although all the variables represented in the figures, i.e. gender, income, education and ethnicity, have been found to be key

Table 2 Quantile regression estimates for selected quantiles of the fruit and vegetable (F&V) intake distribution and multiple linear regression estimates (Ordinary Least Squares, OLS) of the effects of sociodemographic variables on F&V consumption (in 80 g portions/d): data from 11 044 adult individuals in the 2003 Health Survey of England

Variable	Units	0·05 quantile		0·25 quantile		0·50 quantile		0·75 quantile		0·95 quantile		OLS	
		Estimate	P value	Estimate	P value	Estimate	P value	Estimate	P value	Estimate	P value	Estimate	P value
Intercept		−0·05	0·45	1·17	<0·0001	2·29	<0·0001	3·90	<0·0001	7·27	<0·0001	2·74	<0·0001
Equivalised income	1000s of £	0·001	0·20	0·004	0·02	0·005	0·01	0·01	<0·0001	0·01	0·04	0·01	<0·0001
Equivalised income squared	1000s of £	0·00	0·58	0·00	0·29	0·00	0·03	0·00	0·02	0·00	0·01	0·00	0·00
Household size	No. of people	−0·03	0·14	−0·14	0·00	−0·14	0·00	−0·25	<0·0001	−0·68	<0·0001	−0·23	<0·0001
Household size squared	No. of people	0·01	0·31	0·03	0·03	0·03	0·04	0·06	0·00	0·10	0·00	0·04	0·00
Children	No. of children	−0·01	0·64	0·00	0·96	−0·03	0·52	−0·01	0·79	0·28	0·00	0·04	0·31
Age	Years	0·01	<0·0001	0·03	<0·0001	0·03	<0·0001	0·03	<0·0001	0·03	<0·0001	0·03	<0·0001
Age squared	Years	0·00	0·07	0·00	0·87	0·00	0·16	0·00	0·00	0·00	<0·0001	0·00	0·00
Men	Dummy	−0·11	<0·0001	−0·33	<0·0001	−0·43	<0·0001	−0·46	<0·0001	−0·31	0·02	−0·37	<0·0001
Single	Dummy	−0·14	0·00	−0·32	<0·0001	−0·31	<0·0001	−0·29	0·00	−0·23	0·16	−0·27	<0·0001
Degree	Dummy	0·77	<0·0001	1·11	<0·0001	1·33	<0·0001	1·28	<0·0001	1·86	<0·0001	1·21	<0·0001
A-levels	Dummy	0·28	<0·0001	0·72	<0·0001	0·94	<0·0001	0·93	<0·0001	1·44	<0·0001	0·86	<0·0001
Up to GCSE	Dummy	0·17	<0·0001	0·34	<0·0001	0·47	<0·0001	0·40	<0·0001	0·45	0·00	0·41	<0·0001
Managerial class	Dummy	0·13	0·00	0·40	<0·0001	0·42	<0·0001	0·48	<0·0001	0·47	0·00	0·40	<0·0001
Intermediate class	Dummy	0·05	0·17	0·16	0·01	0·23	0·00	0·23	0·01	0·19	0·23	0·16	0·01
Mixed ethnicity	Dummy	0·04	0·86	0·49	<0·0001	0·77	0·00	0·70	0·11	1·59	0·02	0·73	0·00
Black ethnicity	Dummy	0·24	0·29	0·25	0·01	0·41	0·02	0·14	0·33	0·05	0·85	0·40	0·03
Asian ethnicity	Dummy	0·15	0·01	0·61	<0·0001	0·87	<0·0001	0·96	<0·0001	1·48	<0·0001	0·90	<0·0001
North	Dummy	0·07	0·14	0·02	0·78	−0·03	0·80	0·01	0·93	−0·25	0·23	0·03	0·77
West Midlands	Dummy	0·12	0·02	0·24	0·01	0·15	0·15	0·17	0·24	−0·18	0·41	0·14	0·18
East	Dummy	0·15	0·00	0·20	0·00	0·26	0·00	0·33	0·00	0·19	0·24	0·28	0·00
South West	Dummy	0·27	0·00	0·20	0·05	0·08	0·44	0·05	0·72	−0·09	0·71	0·12	0·24
Urban	Dummy	−0·10	0·00	−0·14	0·01	−0·08	0·23	−0·09	0·31	−0·10	0·48	−0·08	0·21
Rural	Dummy	0·08	0·04	0·17	0·00	0·24	0·00	0·26	0·00	−0·02	0·88	0·20	0·00
Spring	Dummy	0·07	0·05	0·05	0·36	0·06	0·38	0·10	0·28	−0·13	0·39	0·07	0·23
Summer	Dummy	0·12	0·00	0·12	0·04	0·25	0·00	0·24	0·01	0·16	0·34	0·23	0·00
Winter	Dummy	−0·01	0·86	0·05	0·41	0·09	0·19	0·07	0·45	−0·01	0·97	0·04	0·54
Fair or bad health	Dummy	−0·14	<0·0001	−0·21	<0·0001	−0·27	<0·0001	−0·26	<0·0001	−0·29	0·02	−0·24	<0·0001
BMI	kg/m²	0·00	0·06	0·01	0·00	0·02	0·00	0·02	0·01	0·01	0·34	0·01	0·00

determinants of F&V intake in other studies and have been confirmed so in the present study, the effects of these variables at low levels of intake are generally weak in comparison to the effects of the same variables at higher levels of intake. Interestingly, we found this effect to persist across most of the other variables included in the analysis, as seen in Table 2. Both the OLS and the QR results show that individuals living singly, in comparison to those living with partners, consume less F&V, all else held equal. However, this effect is lowest at the 0·05 quantile where single individuals consume 0·14 portions less, while the effect rises to 0·32 portions at higher quantiles in the distribution, gradually declining to 0·27 portions less at the right tail. Increased household size is found to negatively impact individual F&V intake. The OLS results show that, for every additional person in the household above the median size, there is about a quarter of an F&V portion reduction (after accounting for the quadratic term). However, the QR results show that the effect is almost half that suggested by OLS for the bottom half of the intake distribution. Those involved in 'intermediate' or 'managerial' socio-economic/occupational classes consume more F&V than those in 'routine' class. However again, the effect is relatively muted where baseline intakes are very low and in the case of the intermediate class, is statistically insignificant at the lower extreme.

Conclusions

Quantile regression methods have much to offer the investigation of the determinants of dietary intake. Dietary inadequacy or excess occurs at the tails of nutrient and food intakes, and it seems intuitive that intake responses in these areas will differ from elsewhere along the intake distribution. Our application of regression methods to F&V intake in England confirms that, even after controlling for a range of variables, we can conclude that women consume more F&V than men; Asians, particularly, but also blacks, more than whites; co-habiting individuals more than single-living ones; and rural inhabitants more than suburban and urban. Increased incomes and education, and reduced household sizes, also boost F&V intake, although the income effect appears small when other factors are controlled for. However, the key general finding of the present study is that the influence of most sociodemographic variables on F&V consumption in England is relatively weak in the area of greatest concern, i.e. at the lower tail where intake inadequacy is severe. One interpretation of this is that those in the lower tail of the conditional distribution (i.e. poor performers within reference groups defined by specific values for socio-economic variables) have inherent traits/preferences, unrelated to any particular socio-economic configuration, that cause them to be poor F&V consumers. This is worrisome from the point of view

of F&V programmes and policies, since it implies that there are few identifiable levers or easy targets when it comes to shrinking the lower tail that represents gross intake inadequacy. Those with grossly inadequate F&V intake are naturally the ones that the 5-a-day campaign would wish to influence with the greatest urgency. However, given that our results find few socio-economic levers influencing such poor performers within any reference group, campaigns would have to be broad-based rather than finely targeted to effect significant improvements among those with very low intakes. This would mean campaigns spanning the spectrum of geographical areas and social and economic classes, and would inevitably require larger budgets.

Acknowledgements

References

1. Doyle M & Hosfield N (2003) *Health Survey for England 2001: Fruit and Vegetable Consumption.* London: The Stationery Office.
2. Leather S (1995) Fruit and vegetables: consumption patterns and health consequences. *Br Food J* **97**, 10–17.
3. Pollard J, Greenwood D, Kirk S & Cade J (2002) Motivations for fruit and vegetable consumption in the UK Women's Cohort Study. *Public Health Nutr* **5**, 479–486.
4. Subar AF, Heimendinger J, Patterson B, Krebs-Smith S, Pivonk E & Kessler R (1995) Fruit and vegetable intake in the United States: the baseline survey of the Five A Day for Better Health Program. *Am J Health Promot* **9**, 352–360.
5. Wardle J, Parmenter K & Waller J (2000) Nutrition knowledge and food intake. *Appetite* **34**, 269–275.
6. Pollard J, Greenwood D, Kirk S & Cade J (2001) Lifestyle factors affecting fruit and vegetable consumption in the UK Women's Cohort Study. *Appetite* **37**, 71–79.
7. Marmot MG (1998) Improvement of social environment to improve health. Viewpoint. *Lancet* **351**, 57–60.
8. Koenker R & Bassett GS (1978) Regression quantiles. *Econometrica* **46**, 33–50.
9. Variyam JN, Blaylock J & Smallwood D (2002) Characterizing the distribution of macronutrient intake among US adults: a quantile regression approach. *Am J Agric Econ* **84**, 454–466.
10. Gustavsen GW & Rickertsen K (2006) A censored quantile regression analysis of vegetable demand: the effects of changes in prices and total expenditure. *Can J Agric Econ* **54**, 631–645.
11. Koenker R & Hallock K (2001) Quantile regression. *J Econ Perspect* **15**, 143–156.
12. Cade BS & Noon BR (2003) A gentle introduction to quantile regression for ecologists. *Front Ecol Environ* **1**, 412–420.

13. National Centre for Social Research & University College London, Department of Epidemiology and Public Health (2005) *Health Survey for England, 2003* (computer file). Colchester: UK Data Archive. SN: 5098.
14. Blake M, Devrill C, Prescott A, Primatesta P & Stamatakis E (2004) *Health Survey for England 2003: Methodology and Documentation*. London: The Stationery Office.
15. He X & Hu F (2002) Markov chain marginal bootstrap. *J Am Stat Assoc* **97**, 783–795.
16. Dibsdall LA, Lambert N, Bobbin RF & Frewer LJ (2003) Low-income consumers' attitudes and behaviour towards access, availability and motivation to eat fruit and vegetables. *Public Health Nutr* **6**, 159–168.
17. Ricciuto L, Tarasuk V & Yatchew A (2006) Socio-demographic influences on food purchasing among Canadian households. *Eur J Clin Nutr* **60**, 778–790.
18. Giskes K, Turrell G, van Lenthe FJ, Brug J & Mackenbach JP (2006) A multilevel study of socio-economic inequalities in food choice behaviour and dietary intake among the Dutch population: the GLOBE study. *Public Health Nutr* **9**, 75–83.
19. British Hearth Foundation (2008) Ethnic Differences in Diet. http://www.heartstats.org/datapage.asp?id=932 (accessed February 2009).

Appendix

More on quantile regression methods

Koenker and Bassett[8] noted that a set of causal variables could have myriad effects on the distribution of the dependent variable. They proposed that conditional quantiles of the dependent variable be estimated as linear functions of covariates, whereas simple linear regression expresses only the conditional mean as a linear function of covariates. By allowing conditional functions to be defined at any chosen quantile, QR leads allows the effect of a given set of covariates to flexibly vary across the distribution of the dependent variable. Unlike the standard applications of logistic regression methods on intake data, no sacrifice of information is entailed, while assumptions about the form of the parametric distribution (such as the logistic distribution in logistic regressions) are avoided. Classical linear regression reduces to a special case of QR where the effects of covariates are constrained to be the same across the distribution of the dependent variable. As Koenker and Hallock[11] caution, simply dividing the data into subsets based on values of the dependent variable and applying linear regression to the subsets is not statistically appropriate, and not comparable to quantile regressions. QR fitting at any quantile incorporates information from all sample data points.