The Use of Scanner Data for Research Into Nutrition

RACHEL GRIFFITH† AND MARTIN O'CONNELL‡

†*Institute for Fiscal Studies and University College London*
‡*Institute for Fiscal Studies*

Abstract

Data from market research firms are increasingly being used by social science researchers. These data provide potentially useful information, including detailed nutritional information and well measured prices, and their panel structure is appealing as it permits researchers to control for unobservable time-invariant household characteristics and model dynamic aspects of household behaviour. We summarise the information on the nutrients in foods that is contained in one source of market research data. We show that there is a lot of variation in nutrients at the individual product level, even within narrowly defined food categories such as butter. We also show that the duration of time over which data is collected can have important implications for analysis of household level nutrient purchases.

1

I. Introduction

Data from market research firms are increasingly being used by social science researchers.[1] In the UK one source of these data is from the market research firm Taylor Nelson Sofres (TNS). There are a number of features of the TNS data that make them well suited for research into how consumers behave with respect to their food purchases. In this article we focus on one particular aspect of the data – the information on the nutrient value of food products. We describe what is contained in these data and some patterns in the nutrients in food purchased and brought into the home. We also make some comparisons to nutrient data from the Expenditure and Food Survey (EFS), which is a well established data source that is used widely by social scientists to analyse consumer behaviour and nutrition in the UK.

We show, firstly, that the very disaggregate nutrient information provided in the TNS data reveal a lot of variation in the nutrient content of food products that are often grouped into the same broad food category. In other data sources, like the EFS, one nutrient value is given for each food category – equal to the weighted average of the nutrients in products within that category. This can mask large differences in the nutrient content of products grouped into the same food category. Secondly, we show that households in the TNS data, on average, record purchasing less energy than those in the EFS. This underreporting varies with household composition, with larger households and households including children reporting relatively less in the TNS. However, ignoring weeks where households participating in the TNS record purchasing no products eliminates most of the gap, suggesting that the main source of the discrepancy rests with how we treat periods of non-reporting. Thirdly, we highlight how the panel structure of the TNS data, and in particular, how it allows us to observe households' purchasing behaviour over relatively long time periods, can lead to improved measurements of the nutrient purchasing behaviour of households. Specifically, we show that

---

[1] See, inter alia, Feenstra and Shapiro (2003), Aguiar and Hurst (2007), Blow, Browning and Crawford (2008), Etile (2008), Griffith et al (2009), Einav et al (2008).

2

using a longer duration of purchases, lowers the dispersion in the distribution of energy obtained from different macronutrients across households. We also show that using longer durations can alter conclusions about the average amount of energy households obtain from macronutrients. Finally, we illustrate how this may lead us to make inaccurate conclusions about which households' purchasing behaviour accords with government recommendations.

There is a large literature on the economic aspects of nutrition in developing countries. More recently, economists have become increasingly interested in nutrition in developed countries, and there is a rapidly growing literature that attempts to understand the economic determinants of nutritional status and to evaluate potential policy interventions. A few examples of issues that have been considered include: the possible links between food prices, time use and obesity (Cutler, Glaeser, and Shapiro (2003), Lakdawalla, Philipson, Bhattacharya (2005) and Bertrand and Schanzenbach (2009)), the impact of economic downturns on nutrition (Stillman and Thomas (2008)), the potential impacts of levying a tax on fats in food (Leicester and Windmeijer (2004), Chouinard, Davis, LaFrance and Perloff (2007) and Griffith, Nesheim and O'Connell (2009)) and from the statistics literature, how we can use data measured at the household level to make statements about individual outcomes (Chesher (1997, 1998)). These are only a few examples of the many potentially interesting areas that these data allow us to explore.

In the next section we briefly describe what is in the data. In section three we discuss the information that is available on nutrients and how it compares to similar data in the EFS. We highlight the large variation in nutrients in products within food categories observed in the TNS data. In section four we consider measures of household level nutrient purchases using the TNS data and compare these to government guideline levels. We show that using a longer duration of purchases improves measures of households' nutrient purchasing behaviour and that using short durations can lead us to make

3

incorrect conclusions about how household behaviour compares with government guidelines. In section five we briefly discuss prices in the TNS data. A final section summarises.

II. Data

The data we use in this paper are from the Taylor Nelson Sofres (TNS) World Panel for 2006 and from the Expenditure and Food Survey (EFS) for 2005-06. Both of these are described in detail in Leicester and Oldfield (this volume), here we summarise the key features of these data, and give more detail on the nutritional information they contain.

1. TNS World Panel

The TNS World Panel contains information on food and drink purchases brought into the home. Participating households use handheld scanners in the home to scan the barcodes of all purchases brought into the home. This information is sent electronically to TNS, along with the till receipts. The data contain precise details on the products that a household purchased, including the purchase price and the date and store of purchase.

The data are longitudinal – households are recruited into the survey and continue to participate for potentially many years. This allows us to observe the purchasing behaviour of the same households for a relatively long duration. The data include information on household characteristics. These are collected annually via a questionnaire and include information on the age and sex of all household members, household income and a number of other household characteristics.

In calendar year 2006 we observe 25,920 households containing 70,315 individuals. We do not observe all households throughout the whole of calendar year 2006. Table 1 shows the duration for which we observe households, broken down by household composition. We observe 15,282 households (around 60%) participating throughout the entire calendar year.

4

The composition of households that remain in the TNS sample for a year is not very different from those that are in it for shorter durations (comparing the first column to the sixth column).

TNS also collect information on product characteristics. Since 2006 these characteristics have included information on the energy, fat, protein, carbohydrates, sodium, sugar and fibre content of food products. TNS collect this information from a range of sources, including manufacturer databases and the nutritional labels on food packaging, and they also test products directly. For most products the information on nutrients is specific to that product, for example, one brand of butter might have a different recorded level of saturated fat than another brand. For some products, such as those with very small market shares, the data on nutrients may be an estimate or average value.

For our analysis we use data on all food and drink (excluding alcohol) purchases. Around 0.7% of observed purchases contain missing nutrient or pack size data. After dropping these observations, our data consist of information on over 23 million purchases of over 100,000 separate products.

2. Expenditure and Food Survey

We compare the TNS data to the EFS data for 2005-06, available from the ESRC Data Archive. The EFS now incorporates what used to be the National Food Survey (NFS), and is one of the most commonly used sources of information on nutrition in social science research in the UK. The EFS data for 2005-06 include information from 16,085 individuals in 6,746 households. The final column of Table 1 shows the number of households of different compositions. The EFS is a diary based survey. All individuals aged 16 and over keep a diary record of daily expenditure for a two-week period. Children keep simplified diaries. We use the data on food and drink (excluding alcohol) expenditures. Data are available to researchers at the level of roughly 500 food and drink categories – 230 of which relate to food

5

brought into the home.[2] The diaries are expenditure based, but quantities are also recorded where possible. If quantities are not recorded then they are estimated using standard portion sizes. Data are available on food purchased and brought into the home, including expenditure and quantity for each household on 230 food categories, and associated data on the nutrient value of each of these 230 food categories.[3]

The Department for Environment, Food and Rural Affairs (DEFRA) estimate quantity, energy and nutrient conversions using the EFS data relating to food purchases.[4] DEFRA's estimates of nutrient intakes are calculated using nutrient composition data supplied by the Food Standards Agency (FSA). The majority of the data is from the FSA's nutrient analysis programme, supplemented by values from manufacturers and retailers. Each of the 500 food codes in the EFS is made up of a number of sub-codes with nutrient composition data attached. A weighted average nutrient composition is calculated for each food code based on estimates of the market share of each sub-code. The nutrient composition data is updated quarterly to keep information in line with new or reformulated products. How important is this variation in nutrients over time? Table 2 summarises the amount of variation in nutritional values that we see over a one year period for the 230 food categories that are associated with food brought into the home. The table shows that there is very little variation over time, with 97% of food categories never changing and a small number of categories seeing relatively minor changes to their nutrient composition. Thus it seems that, at least for these macronutrients, the variation over short time periods is relatively unimportant.

Before comparing the information in the two datasets it is worth commenting on the relative importance of food brought into the home and

---

[2] The other 270 categories relate to food not taken into the home and take-away food. Information on food not taken into the home and purchases of take-away food is not included in the TNS data that we use, and therefore we discard these purchases when using the EFS data.

[3] The EFS contains information on a much larger set of nutrients than the TNS, including not only macronutrients but also vitamins and minerals.

[4] Further information at: https://statistics.defra.gov.uk/esg/publications/efs/method/default.asp

food eaten out. Food eaten out accounts for on average 32% of households' total food expenditure, but only 12% of energy purchased. This is because food eaten out is more expensive per unit of energy – the price includes preparation and service as well as the cost of the food ingredients. Column 1 of Table 3 shows the proportion of expenditure on food eaten out by household composition. Households with no children spend more on average on food eaten out, relative to households with the same number of adults, but with children. However, households with no children get less of their total energy from food eaten out, as shown by column 2. Overall, Table 3 shows that the relative importance of food eaten out is much less with respect to energy purchased than with expenditure.

III. Comparison of nutrients at the product level

We start by describing the variation in macronutrients (carbohydrates, protein and fat – which we separate into saturated fat and unsaturated fat) between individual food products that are grouped within the same food category in the TNS data, and compare that to the information available in the EFS. We see that there is substantial variation across food products within the same food category, meaning that individual households may be purchasing very different amounts of macronutrients than is implied by the aggregated nutrient values reported in the EFS.

1. Within food category variation in nutrients

In data collected by market research firms such as TNS the information on purchases are available at the level of the specific product (the barcode level), and thus they provide very precise information on product characteristics. In contrast, the EFS provides data on nutrients at the more aggregated level of 230 food categories. How much difference this makes in practice will depend on how much variation there is in nutrients within each

food category. Our results suggest that aggregation masks considerable variation in nutrient content across specific products.

To illustrate we select four food categories – butter, yoghurt, minced beef and peanut butter – and consider four examples of variation within each food category, one for each macronutrient.[5]

Variation in saturated fat within butter products

In the TNS information on each macronutrient is available at the barcode level, so we have a different value of each macronutrient for each different butter product. Figure 1 shows the distribution of the intensity of saturated fat (defined as the amount of saturated fat per 100g of product) across all butter products that we observe purchased in the TNS data in 2006, weighted by the number of purchases observed. In the EFS one value is provided for the saturated fat intensity for the food category butter – this is shown by the vertical line at 53.3g. The nutrient value used in the EFS is indeed very close to the modal value observed in the TNS purchase data, however it is clear that there is substantial variation around this, with, in particular, some households purchasing butters that are much lower in saturated fat than the mode.

Variation in carbohydrates within yoghurt products

As a second example we consider yoghurt. Figure 2 shows the distribution of carbohydrate intensity across all yoghurt product purchases in the TNS data, with the vertical line indicating the aggregated value of 14.6g provided in the EFS. Again we see that this is close to the modal value, but that there is substantial variation around this.

---

[5] We do not have an exact mapping between food categories in the TNS and EFS data, meaning that in general it is difficult to match TNS products into EFS food categories. However, TNS groups products into food categories. While these categories do not in general correspond exactly to the EFS categories, it does allow us to pick some food categories that are comparable.

Variation in protein within minced beef products

Our third example is based on the protein intensity of minced beef products. Figure 3 shows the distribution across all minced beef purchases observed in the TNS data. Once again the EFS value of 20.3g comes close to the modal TNS value across all purchases of the various minced beef products. However, as in the two previous examples the EFS value masks considerable within product category variation.

Variation in unsaturated fat within peanut butter products

As a final example we look at the intensity of unsaturated fat in peanut butter products. Figure 4 shows that this time the EFS value for the unsaturated fat intensity of peanut butter products of 42.0g is higher than the TNS modal value of 40.0g. However, the variation in the macronutrient intensity is considerably lower than in the previous examples – the coefficient of variation is 0.08, compared with 0.41 for the carbohydrate intensity of yoghurt products. In this case, the EFS aggregate intensity provides a reasonably good approximation for the intensities of most products contained in the TNS data.

2. Overall variation within product categories

The examples above illustrate that the variation in nutrients within product categories is potentially important and of interest to researchers investigating the nutritional implications of households' purchasing behaviour. Are these examples typical of other food categories? It is difficult to compare exactly across the two datasets because products are coded differently – the TNS food categories do not exactly correspond with the EFS ones. To provide a rough idea of the full extent of variation within food categories we use the TNS coding that divides products into 193 food categories. These are not the same as the 230 EFS food categories, but provide a rough approximation.

9

In Table 4 we report the mean nutrient intensity and the distribution of the standard deviations of each of the macronutrients, as well as fibre, sodium and sugars within the 193 TNS food categories. For example, consider carbohydrates in the first row. The mean intensity of carbohydrates across all purchases is 21.31g. The mean standard deviation of 6.48g tells us that on average, across all 193 food categories, approximately two-thirds of purchases of products from a given food category are within 6g of the mean carbohydrate intensity for their category (while one-third are either 6g more or less carbohydrate intensive). The median is somewhat lower at 4.54g. The minimum is zero; there are four food categories in TNS in which there is no variation in the recorded carbohydrate intensity – hens' eggs, frozen Christmas puddings, pepper, herbs and spices, and frozen stuffing.[6] The maximum is 32.42g; this is for the food category milkshake mixes, where there is a lot of variation in the carbohydrate intensity of different products within the group.

Overall Table 4 shows that there is a lot of variation in the nutrient values of specific products within food categories. This may well be important for any analysis of household nutrient purchasing behaviour.

The more accurate nutrient information provided in the TNS data are useful as they allow researchers to gain a more accurate picture of the nutrients purchased by individual households. For example, a household that purchased Anchor Lighter butter purchased a product with 23.7g of saturated fat per 100g of product. If the household was observed in the TNS data, we would have accurate information on this purchase. If, on the other hand, we observed the household in the EFS data, we would observe it purchasing butter which has an associated saturated fat intensity of 53.3g – over double the true value. Clearly the information provided in the TNS helps us more accurately measure the nutrient purchasing behaviour of households.

---

[6] The absence of variation may reflect an actual lack of variation, or it may be due to TNS using an average nutrient value for products that are infrequently purchased.

In addition, the more disaggregate nutrient information allow researchers to answer particular research questions that would be more difficult, or impossible, to address with aggregated data. For example, Griffith, Nesheim and O'Connell (2009) use the TNS data to investigate the effectiveness of imposing a tax on saturated fat. One potentially important margin of consumer response to this policy is substitution to similar products within the same food category. Since the TNS data provide nutrient information at the barcode level the authors can simulate the impact of the policy allowing for the differential price increases the tax would imply, for example, for different butter products. This allows them to model consumers' willingness to switch to butter products which contain less saturated fat. If instead they used the EFS data, they would be unable to model within food category substitution, meaning, for instance, that in their model households would be constrained to either adjust the quantity of butter they purchase or to switch entirely to another food category.

IV. Comparison of data on nutrients at the household level

In this section we show that the total energy that households in the TNS report purchasing tends to be lower than households in the EFS. This mirrors the description in Leicester and Oldfield, where they show that households in the TNS report lower levels of expenditure. However, the differences between the TNS and EFS are substantially reduced if we focus on periods where households consistently record making purchases. Periods where zero purchases are recorded might reflect genuine gaps in expenditure, or they might represent underreporting as households experience reporting fatigue.

In what follows we focus our analysis on the proportion of energy that households purchase in the form of each macronutrient. We show that the duration of the sample period has important implications for the measurement of these proportions. Shorter durations lead to systematic mismeasurement of nutrient purchases, as well as to more noisy measures. We then compare these measures to government consumption guidelines.

11

1 Energy purchased

The total amount of energy that households report purchasing in the TNS data is typically lower than that reported by households in the EFS. This is likely to be due, at least in part, to underreporting, as discussed in Leicester and Oldfield (this volume). Table 5 shows the average amount of energy purchased as food and drink (excluding alcohol) that was taken into the home per household member per day. The rows show the average for different household compositions. The first four columns use the TNS data and report the average using different samples of household and different durations of participation, while the last column uses the EFS data.

Column one is based on all households and all observed purchases in the TNS data. We calculate the average energy purchased per person by summing up all energy purchased over the household's period of participation in the survey and dividing by the number of people in the household and the number of days of participation. The averages in the second column are based only on households that participated throughout the entire year, and are calculated using the entire 52-week duration. Columns 3 and 4 replicate columns 1 and 2, but instead of using all observed purchases, they are based on the longest period of uninterrupted purchases for each household. This means that only purchases made in the longest period where the household did not have a week where it recorded making zero purchases are used. Therefore, weeks with zero recorded expenditure are omitted. For the sample of all households the mean longest period of uninterrupted purchases across households is 15.7 weeks, while for those households that participated for the entire 52 weeks, it is 19.9 weeks. The value is larger for the households who participated throughout 2006, in part, because they were in the sample for longest. The final column shows the average amount of energy purchased as food and drink (excluding alcohol) that was taken into the home per household member per day by households in the EFS data. Table 1, columns (1), (6) and (7) show the number of households in each category.

12

Comparing the first and final columns we see that households in the TNS data record purchasing less energy per household member per day than those in the EFS data, and that this underreporting of energy purchases varies systematically with household composition. In particular, the difference between the TNS and EFS is largest for households with more people, and especially with more children.

Column two shows that conditioning on households that participated throughout the year increases the average amount of energy per household member per day that was reported purchased by between 26 kilocalories for households with two adults and at least four children to 96 kilocalories for households with one adult and one child. Columns three and four show that focusing on the longest period of uninterrupted purchases for each household significantly reduces the gap between energy recorded purchased by households in the TNS data and those in the EFS. In fact, for one adult households the gap is completely eliminated, although the gap is not completely closed for larger households. The difference between columns one and two on the one hand, and three, four and five on the other, suggests that the treatment of periods during which there are zero purchases is one of the reasons that differences arise between the TNS and EFS. For any particular application it may be important to understand more about why these arise and how they should be treated.

2 Duration of data collection

We now focus on the proportion of households' total energy purchases that come from each macronutrient. The TNS data include information on households' purchases over a long duration of time (in some cases several years), whereas for each household the EFS contains information on food purchases over a 2 week duration. In this section we show that this has potentially important implications for the measurement of nutrient purchases.

13

Impact of duration on estimation of the proportion of macronutrients purchased

Using a longer duration of data on purchases allow us to obtain a less noisy estimate of the proportion of energy a household obtains from each macronutrient.[7] We also show that there is a potential bias in the estimated proportion of energy from some macronutrients when we use shorter durations of purchases. This arises because households may purchase systematically different products in smaller or larger shopping trips (or over periods of time when they purchase a little or a lot), and sampling over, say, a 2 week duration compared to over a 52 week duration will weight these different trips differently. The sample that uses a 2 week duration gives equal weight to each 2 week period, since each period has an equal chance of being sampled. Whereas the sample that uses a 52 week duration weights each two week period according to the amount of energy purchased in that two weeks.

Consider a stylised example. Suppose that a household undertakes two types of shopping trips – a main trip and a top up shop. The main trip accounts for 75% of energy purchased, while the top up shop accounts for 25%. The composition of purchases varies systematically between these trips, with more of certain types of foods purchased on the main trips (those that are storable, difficult to transport, etc.). Let $e(n)_s$ denote the amount of energy obtained from macronutrient *n* in trip type *s*, where *n*=carbohydrate, protein, fat. Then the proportion of energy that is purchased in trip type *s* in the form of a macronutrient *n*, which we denote $p(n)_s$, is:

$$p(n)_s = \frac{e(n)_s}{\sum_n e(n)_s}$$

Assume that the proportions of energy purchased in the form of carbohydrate is 45% on the main trip and 65% on the top up trip, so the overall proportion of energy purchased as carbohydrates is 50% (i.e. 75%*45% + 25%*65%).

---

[7] Ahmed et al (2006) compare the variance in diary and recall measures of consumption.

To keep the example simple, assume that in any 2 week duration we only observe one trip per household, and that we observe either a main ($s=1$) or top up ($s=2$) trip with equal probability.[8] Consider the measured proportion of energy from carbohydrate, p(c). For any household we observe only one trip so that trip will comprise all of the energy purchased. When we take a sample we are equally likely to observe either a main trip or a top up trip, so the expected value of p(c) is an average of the values observed on these trips, and is given by:

$$E[p(c)] = 0.5p(c)_1 + 0.5p(c)_2$$
$$= 0.5 * 45\% + 0.5 * 65\% = 55\%$$

In contrast, if we sample a longer duration where we observe both main and top up trips, this means that the probability of observing each type of trip is now one, and that the proportions will be weighted by the share of total energy purchased on that trip. In this case the expected value of p(c) that we calculate reflects the fact that more energy is purchased on main trips (where foods that have relatively less energy as carbohydrates are purchased), and less energy is purchased in top up trips (where food that are relatively carbohydrate intensive are purchased). This is given by:[9]

$$E[p(c)] = 0.75p(c)_1 + 0.25p(c)_2$$
$$= 0.75 * 45\% + 0.25 * 65\% = 50\%$$

The bias arises because using a shorter duration means we weight each type of shopping trip incorrectly. In this case the weight is equal to the probability of observing the shopping trip. When we use the longer duration, we observe each shopping trip with certainty, meaning each trip is weighted by how much energy, relative to the total, that is purchased in that trip. If in each type of trip, the households obtained the same proportion of their energy as carbohydrate, then the two durations would yield the same answer,

---

[8] We can write this more generally to allow for many different types of shopping trips (main, top up, birthdays, bank holidays, Christmas, etc.).

[9] This can be seen by writing the expected value of the proportion of energy from carbohydrates as $E[p(c)] = (e(c)_1 + e(c)_2)/(\sum_n e(n)_1 + \sum_n e(n)_2)$, noting that this can be rewritten as $E[p(c)] = (\sum_n e(n)_1/(\sum_n e(n)_1 + \sum_n e(n)_2))*p(c)_1 + (\sum_n e(n)_2/(\sum_n e(n)_1 + \sum_n e(n)_2))*p(c)_2$, and noting that from the example $(\sum_n e(n)_1/(\sum_n e(n)_1 + \sum_n e(n)_2)) = 0.75$ and $(\sum_n e(n)_2/(\sum_n e(n)_1 + \sum_n e(n)_2)) = 0.25$.

despite the 2 week duration using incorrect weights. However, in this example – and, as we will see, in our data – the proportion of energy bought in any shopping trip is correlated with the proportion of energy obtained as a given macronutrient. It is this combination of incorrect weights and correlation between the amount of energy purchased and the amount of it purchased as a given macronutrient that gives rise to the bias associated with the shorter duration.

If the probability that we observe a particular trip in any two week period is not equal to one and if the total proportion of energy purchased in that kind of trip, is correlated with the proportion of energy purchased on that trip as a particular macronutrient then this leads to bias in the estimate. To get some idea of whether this is true we calculate the correlation between the proportion of total energy purchased by each household in a given week and the proportion of the energy purchased as a particular macronutrient. This is reported in Table 6. The correlations suggest that in weeks where households buy a relatively large amount of energy they are likely to purchase a relatively large proportion of it as unsaturated fat (and to a lesser extent saturated fat) and a relatively small amount as protein and carbohydrates. This suggests that looking at shorter durations may lead to some bias in the share of energy purchased as specific macronutrients - extending the duration is likely to increase the estimated mean proportion of energy obtained from fat and reduce the estimated mean proportion obtained from protein and carbohydrates.

Description of variation in proportion of macronutrients purchased by duration

To investigate whether the duration matters in practice we use the TNS data to compare the proportion of energy purchased in the form of each macronutrient across households, using data for different durations of participation from 2 weeks to 52 weeks. In all cases except the 52 week duration we select a random period of the given duration in 2006 for each household.

16

Figures 5 – 8 show the distribution across households of energy obtained from each macronutrient. They are based on all recorded purchases of food and non-alcoholic beverages for the sample of households that participated in the survey throughout 2006 (15,282 households), so that each distribution is based on the same sample of households, with the only difference being the duration of data that is used. This avoids differences in the distributions arising through changes in the composition of households included. Each graph contains six distributions – one for each duration.

As described above, we see in Figures 5 – 8 that using longer durations allows us to more precisely measure the nutrients that households purchase – there is lower dispersion across households in the distributions as the duration increases. In each case the standard deviation falls substantially when we move from the 2 week to the 52 week duration.

In addition, we see that there is a shift in the means of the distributions. In Figure 5 we see that the distribution of carbohydrates and protein shifts to the left, suggesting that we overestimate the proportion of energy purchased in these forms when we use shorter durations, while the opposite is true for both unsaturated and saturated fats. This is consistent with our prediction based on the correlations between energy purchased and the proportion of energy purchased as the four macronutrients. In fact, the mean proportions from unsaturated and saturated fat increase by 4% and 1% respectively, while the mean proportions from protein and carbohydrates fall by 2% and 1%.

3 Comparison to recommended levels

One use of nutrition information at the household level is that it can be used to compare household purchasing behaviour to government consumption guidelines. This allows us to make inferences about which kinds of households are most likely to consume excessive or insufficient quantities of nutrients. In this section we use the TNS data to compare the composition of households' energy purchases with the Dietary Reference

17

Values (DRVs) provided by the Department of Health.[10] The DRVs state that the average percentage of energy consumed in the form of each macronutrient that is consistent with good health is 50% for carbohydrates, 15% for protein, 24% for unsaturated fat and 11% for saturated fat.[11]

The previous section suggests that analysis based on comparing household behaviour to DVRs may be affected by the duration of purchases used. We compare both the 2 week and 52 week durations with the DRVs to see the extent to which this is true.

Figure 9 shows a scatter plot of the proportion of energy purchased as carbohydrates across households. The horizontal axis shows the proportion of energy purchased as carbohydrate based on a 2 week duration and the vertical axis shows the proportion of energy purchased as carbohydrate based on a 52 week duration. Each dot represents one of the 15,282 individual households, and it shows the proportion of energy that the household reported purchasing as carbohydrate using the two alternative durations.

If the two measures were identical then we would see a 45° straight line starting with the household that purchased the lowest proportion extending through to the household that purchased the highest proportion. The cloud pattern is a result of the differences in the mean and dispersion of the two measures, shown in Figure 5.

The vertical and horizontal lines correspond to the DRVs (these are the same). These lines split the figure into four quadrants. The percentage of households in each quadrant is also displayed in the figure. The 26% of households that are in the north-east quadrant are ones that purchase more of their energy as carbohydrate than the guideline value using either duration, while the 45% in the south-west quadrant purchase less using data of either duration. The 10% of households in the north-west quadrant purchase more

---

[10] See Report to the Panel on Dietary Reference Values of the Committee on Medical Aspects of Food Poilcy (1991)

[11] This assumes no energy is consumed as alcohol. For those individuals that consume alcohol the Department of Health assign an energy contribution of 5% to alcohol.

than the guideline according to the 52 week duration and less according to the 2 week duration. Conversely, the 19% of households in the south-east quadrant purchase too much energy as carbohydrate using the 2 week duration and too little using the 52 week duration. Figures 10 – 12 replicate the scatter plot for the other macronutrients.

A number of points emerge from these figures. First, Figure 12 suggests that most households (around 95% of all households according to the 52 week duration) purchase more energy in the form of saturated fat than the DRV, and correspondingly, Figures 10 – 11 suggest most households purchase less energy in the form of the other macronutrients than is recommended. Second, there is a lot of heterogeneity across households, with some households purchasing a lot more of a given macronutrient than the DRV, and others purchasing a lot less.

Third, the duration of purchases used to measure the proportion of energy coming from each macronutrient has an important effect on the conclusions that we reach about which households purchase more and which purchase less than the DRVs – i.e. there are many households in the north-east and south-west quadrants – and by how much they do so.

From a preliminary look there does not appear to be any clear pattern between household composition and whether a household purchases more or less than the DRVs. However, this topic merits further investigation.

An important determinant of the difference between a household's proportion using the 2 week duration and the proportion using the 52 week duration is the expenditure the household makes in the 2 weeks that it is sampled. In particular, controlling for household composition, the more a household spends in the 2 week period it is sampled, the closer the proportion based on the 2 week duration is to the one based on the 52 week duration. This makes sense, given that it is more likely that a household will purchase goods that are unrepresentative of its annual energy consumption in a 2 week period where it does not buy very much.

19

V Prices

It is worth making a brief comment on prices. Well measured prices allow researchers to model consumer decision making taking account of variation in the prices of different products. This is important if a researcher wants to quantify how consumers respond to relative price changes, or to assess the impacts of policies that work through the price mechanism (whether they change prices directly or indirectly through influencing firms' pricing decisions). Since we believe that the price of a product is a key factor that influences a household's decision to purchase, it is important to control for it even when analysing consumer behaviour with respect to other product characteristics, or when analysing a policy where the price mechanism is not the main object of interest. For example, it is difficult to assess the impact of the 5-a-day information campaign, where the government tried to encourage people to eat at least five portions of fruit and vegetables each day, without controlling for changes in the relative price of fruit and vegetables that occurred over the duration of the campaign.

The TNS data include detailed information on prices paid for specific products, as well as information on the stores that products were purchased in. In the EFS we have measures of expenditure and quantity for food categories. This means that we can measure "unit values", or a weighted average of the prices paid for the food items included in each food category.[12] These "unit values" vary within food category across households. This presumably reflects differences in the products purchased as well as differences in the prices faced by households. The EFS data obscures this last point by aggregating to the food category level and by aggregating across all purchases made by households in their 2 week participation period. It is beyond the scope of this paper to provide a full description of differences in prices in the TNS data and unit values in the EFS, but this would be an interesting area for future work.

---

[12] See, inter alia, Lahatte et al (1998), Deaton (1987, 2006)

VI. Summary and Conclusions

Scanner data from market research firms are increasingly being used by social science researchers. These data open up the possibility of several new areas of research. For example, using similar data from France, Bonnet, Dubois and Réquillart (2008) are exploring consumer behaviour with respect to the consumption of saturated fat. Griffith, Nesheim and O'Connell (2009) are modelling the potential impact of a tax on saturated fats. Further work with these data could consider the impact of other possible policy interventions, and evaluate the impact of past interventions, such as the 5-a-day campaign, as well as many other issues.

These data provide potentially useful information, including detailed nutritional information and well measured prices, and their panel structure is appealing as it permits researchers to control for unobservable time-invariant household characteristics and model dynamic aspects of household behaviour. In the UK one source of these data is from the market research firm TNS. In this article we discussed one particular aspect of these data – the information they contain on the nutrients in foods. We showed that there is substantial variation in the macronutrients contained in similar food products. By providing nutritional information at the product level, the TNS data offer a significant advantage over other commonly used data sources that contain nutrient information at the aggregated food category level. We also showed that households tend to record purchasing less energy in the TNS data, relative to those in the EFS, but this shortfall is reduced when we omit periods where no purchases are recorded. Finally we illustrated that the duration of purchases used to measure purchases can have important consequences for the precision and accuracy of the measurement of households' nutrient purchases.

21

References

Aguiar, M and E Hurst (2007) "Life-cycle Prices and Production" American Economic Review, Vol 97, No. 5 (December 2007), 1533-1559

Ahmed, Naeem, Matthew Brzozowski and Thomas Crossley (2006) "Measurement Errors in Recall Food Consumption Data" *IFS Working Paper* WP06/21

Bertrand, Marianne, and Diane Whitmore Schanzenbach. (2009). "Time Use and Food Consumption." *American Economic Review*, 99(2): 170–76.

Blow, L, M Browning and I Crawford (2008) "Revealed Preference Analysis of Characteristics Models," *Review of Economic Studies*, vol. 75(2), pages 371-389

Bonnet, Celine, Pierre Dubois and Vincent Réquillart (2008), "*The dynamics of satured fat consumption in France"*, Toulouse mimeo

Chesher, A. (1997), "Diet Revealed?: Semiparametric Estimation of Nutrient Intake--Age Relationships" *Journal of the Royal Statistical Society*. Series A (Statistics in Society), Vol. 160, No. 3 (1997), pp. 389-428

Chesher, A. (1998), "Individual demands from household aggregates: time and age variation in the composition of diet", *Journal of Applied Econometrics*, 13(5), 505-524

Chouinard, H., D. Davis, J. LaFrance and J. Perloff (2007), "Fat taxes: big money for small change", *Forum for Health Economics & Policy* Volume 10, Issue 2

Cutler D. M., Glaeser E., Shapiro J. M. (2003), "Why have americans become more obese?", *Journal of Economic Perspectives*, 17(3), 93-118

Deaton, Angus (1987) "Estimation of own and cross-price elasticities from household survey data" *Journal of Econometrics* 36, 7-13

Deaton, Angus (2006) "Prices, Unit Values and Local Measurement Units in Rural Surveys: an Econometric Approach with an Application to

Poverty Measurement in Ethiopia" *Journal of African Economies*, 15: 2 (June), 181-211

Einav, L., E. Liebtag and A. Nevo (2008), "On the accuracy of Nielsen Homescan data", US Department of Agriculture Research Report 69, Washington, DC: USDA (http://www.ers.usda.gov/Publications/ERR69/ERR69.pdf).

Etile, Fabrice (2008) "Food price policies and the distribution of body mass index: Theory and empirical evidence from France" Paris School of Economics Working Paper No 2008-28

Feenstra, R and M Shapiro (2003), *Scanner Data and Price Indexes*, University of Chicago Press: Chicago

Griffith, Rachel, Ephraim Leibtag, Andrew Leicester and Aviv Nevo (2009) "Consumer Shopping Behavior: How Much Do Consumers Save?" *Journal of Economic Perspective*, 23(2): 99–120

Griffith, Nesheim and O'Connell (2009) "Empirical Estimates of a Fat Tax" IFS mimeo

HMSO (1991) Dietary Reference Values for Food Energy and Nutrients for the United Kingdom: Report to the Panel on Dietary Reference Values of the Committee on Medical Aspects of Food Policy, ISBN 0 11 321397 2

Lahatte, Agenor, Ruth Miquel, François Laisney, and Ian Preston (1998) "Demand systems with unit values: A comparison of two specifications" *Economics Letters* Volume 58, Issue 3, 1 March 1998, Pages 281-290

Lakdawalla D., Philipson T., Bhattacharya J. (2005), "Welfare enhancing technological change and the growth of obesity", *American Economic Review*, May 253-257

Leicester, Andrew and Zoe Oldfield (2009) "The use of scanner technology for collecting expenditure data" *Fiscal Studies*, this volume

Leicester, A. and F. Windmeijer (2004), *The "Fat Tax": economic incentives to reduce obesity* IFS Briefing Note No. 49, June 2004.

Office for National Statistics and Department for Environment, Food and Rural Affairs, *Expenditure and Food Survey, 2005-2006* [computer file]. *2nd Edition.* Colchester, Essex: UK Data Archive [distributor], October 2007. SN: 5688

Stillman, S and D Thomas (2008) "Nutritional Status During an Economic Crisis: Evidence from Russia" Economic Journal, Volume 118 Issue 531, p 1385-1417

Table 1: Number of households by duration of participation

	TNS						EFS
	2 weeks	4 weeks	8 weeks	16 weeks	32 weeks	52 weeks	2 weeks
Adult 1 Child 0	4494	4473	4296	3986	3421	2944	1880
	17%	*17%*	*18%*	*18%*	*19%*	*19%*	*28%*
Adult 1 Child 1+	1384	1380	1269	1142	920	696	475
	5%	*5%*	*5%*	*5%*	*5%*	*5%*	*7%*
Adult 2 Child 0	8526	8491	8169	7501	5934	5068	2260
	33%	*33%*	*33%*	*33%*	*32%*	*33%*	*34%*
Adult 2 Child 1	2863	2845	2646	2418	1920	1504	537
	11%	*11%*	*11%*	*11%*	*10%*	*10%*	*8%*
Adult 2 Child 2	3501	3485	3246	2995	2484	2063	613
	14%	*14%*	*13%*	*13%*	*14%*	*13%*	*9%*
Adult 2 Child 3	1246	1239	1163	1083	890	722	197
	5%	*5%*	*5%*	*5%*	*5%*	*5%*	*3%*
Adult 2 Child 4+	473	471	430	397	329	260	68
	2%	*2%*	*2%*	*2%*	*2%*	*2%*	*1%*
Adult 3+ Child 0	2240	2228	2132	1991	1601	1333	473
	9%	*9%*	*9%*	*9%*	*9%*	*9%*	*7%*
Adult 3+ Child 1+	1193	1189	1101	1014	819	692	243
	5%	*5%*	*5%*	*5%*	*4%*	*5%*	*4%*
Total	*25920*	*25801*	*24452*	*22527*	*18318*	*15282*	*6746*
	100%	*100%*	*100%*	*100%*	*100%*	*100%*	*100%*

Notes: The TNS data are for calendar year 2006. Households are selected on the basis of their recorded start up and drop out dates. The EFS data is for 2005-06 and all households participate for two weeks.

1

Table 2: Changes in nutrients over time in the EFS

	No change	Change of less than 0.5 g or kcal	Change of between 0.5-1 g or kcal	Change of more than 1 g or kcal	Total
			Number of food categories in which there was:		
Carbohydrates (g)	224	2	0	4	230
	97.39	*0.87*	*0.00*	*1.74*	*100.00*
Energy (kcal)	224	0	0	6	230
	97.39	*0.00*	*0.00*	*2.61*	*100.00*
Protein (g)	224	6	0	0	230
	97.39	*2.61*	*0.00*	*0.00*	*100.00*
Saturated (g)	224	5	1	0	230
	97.39	*2.17*	*0.43*	*0.00*	*100.00*
Unsaturated (g)	224	5	0	1	230
	97.39	*2.17*	*0.00*	*0.43*	*100.00*

Notes: Number of food categories is in standard type, the percentage of categories is in italics. The nutrient composition of each category is reported quarterly. Those classified as having changed, must therefore have at least two quarters in 2005-06 in which the nutrient in question has a different value.

Table 3: Food out a percentage of total food

Household composition	Percentage of total food expenditure on food eaten out	Percentage of total energy from food eaten out
Adult 1 Child 0	32.1	9.4
Adult 1 Child 1+	24.8	18.6
Adult 2 Child 0	32.3	9.8
Adult 2 Child 1	30.6	11.6
Adult 2 Child 2	29.6	12.4
Adult 2 Child 3	28.8	13.7
Adult 2 Child 4+	22.6	14.6
Adult 3+ Child 0	36.8	12.9
Adult 3+ Child 1 or 2	35.8	14.2
Adult 3+ Child 3+	30.0	14.8
Adult 4+ Child 0	39.7	12.6

Notes: Figures based on EFS data for 2005-06.

Table 4: Variation of macronutrients within 193 TNS food categories

	Mean nutrient intensity	Mean standard deviation	Median standard deviation	Minimum standard deviation	Maximum standard deviation
Carbohydrates (g)	21.31	6.48	4.54	0	32.42
Protein (g)	7.12	2.16	1.63	0	22.32
Saturated fat (g)	3.91	1.84	1.20	0	14.00
Unsaturated fat (g)	6.98	3.13	2.34	0	23.49
Fibre (g)	1.45	0.72	0.54	0	4.30
Sodium (g)	0.41	0.20	0.12	0	4.47
Sugars (g)	10.17	3.77	2.00	0	31.68

Notes: We calculate the standard deviation of each nutrient within each of the 193 TNS food categories. The numbers reported in the table are the mean, median, minimum and maximum of these 193 values.

Table 5: Comparison of energy (in kcal) per household member per day purchased in TNS and the EFS, by household composition

Household composition	TNS			
	Including weeks with no recorded purchases		Based on longest period of uninterrupte purchases	
	All households	Households participating for 52 weeks	All households	Households participating for 5 weeks
Adult 1 Child 0	1759	1850	2284	2256
Adult 1 Child 1+	1068	1164	1646	1595
Adult 2 Child 0	1567	1655	1979	1972
Adult 2 Child 1	1044	1131	1503	1496
Adult 2 Child 2	967	1032	1387	1387
Adult 2 Child 3	908	980	1334	1318
Adult 2 Child 4+	817	843	1239	1192
Adult 3+ Child 0	1236	1280	1595	1550
Adult 3+ Child 1+	999	1048	1383	1351

Notes: We calculate the total amount of energy recorded as being purchased by each household and divide it by the number of days they participate in the survey and by the number of household members. We then calculate the average value in each household composition category and report the value based on alternative household samples in the table. The number of households in each composition category is given in Table 1.

Table 6: Correlation coefficients for the proportion of annual energy per household purchased in a week and the proportion of energy from each macronutrient in that week

|	Proportion of total energy purchased
Proportion of energy purchased as protein	-0.15
Proportion of energy purchased as carbohydrates	-0.11
Proportion of energy purchased as unsaturated fat	0.21
Proportion of energy purchased as saturated fat	0.04

Notes: We calculate the proportion of each household's annual energy purchases made in each week of 2006 and the proportion of that weekly energy that comes from each macronutrient and report the correlation coefficients. The figures are based on the 15,282 households that participated in the TNS for the complete calendar year 2006.

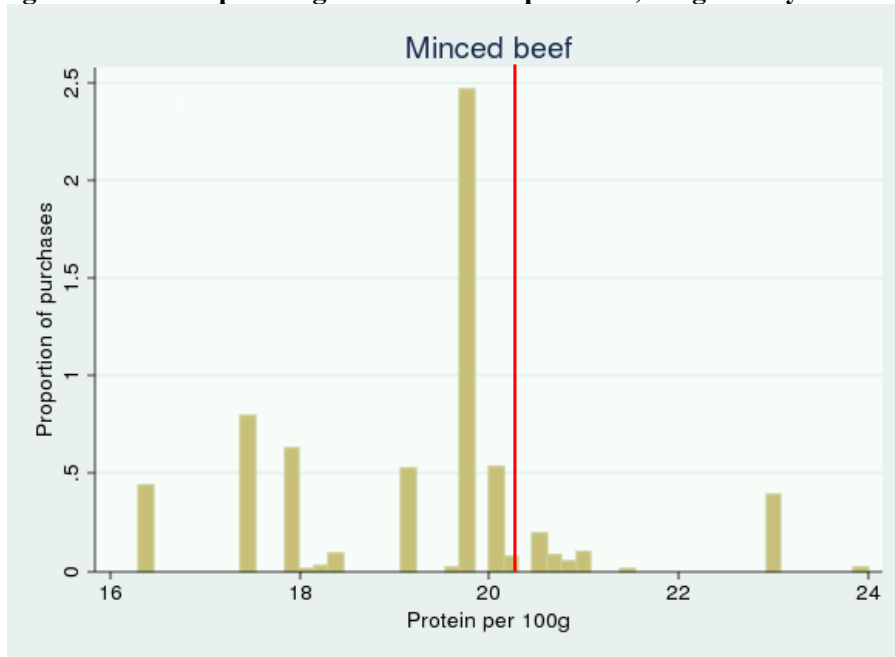**Figure 1: Saturated fat per 100g in butter products, weighted by number of purchases**
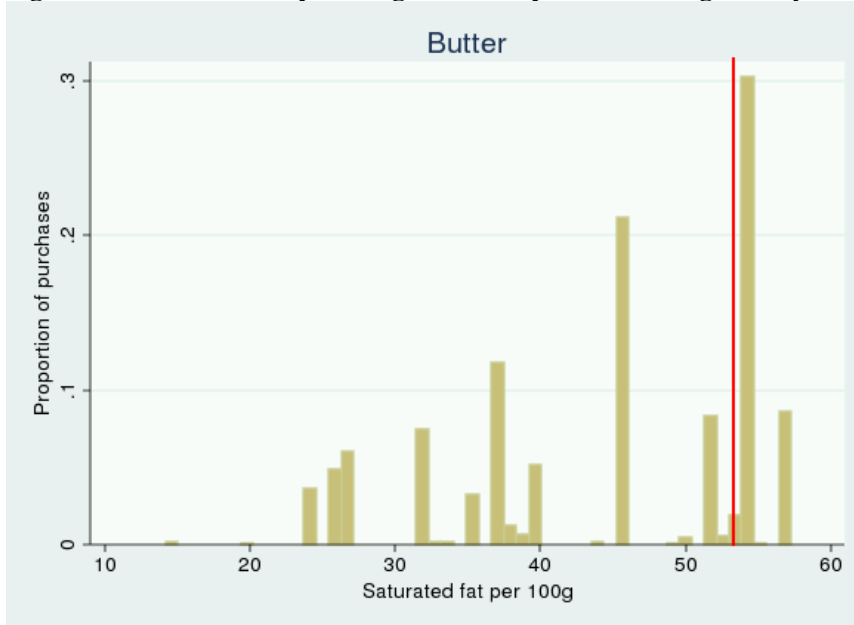
Notes: Figures are grams of saturated fat per 100g of butter. The proportions are based on the number of observed purchases of all butter products in 2006 in the TNS data. The line represents the aggregate EFS intensity 53.3g.

**Figure 2: Carbohydrates per 100g in yoghurt products, weighted by number of purchases**
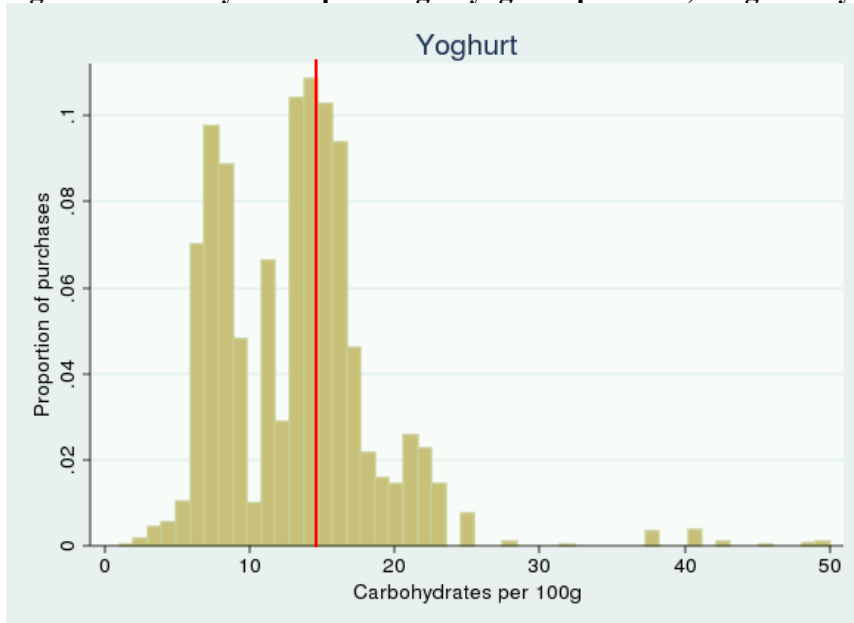
Notes: Figures are grams of carbohydrates per 100g of yoghurt. The proportions are based on the number of observed purchases of all yoghurt products in 2006 in the TNS data. The line represents the aggregate EFS intensity of 14.6g.

Figure 3: Protein per 100g in minced beef products, weighted by number of purchases

Notes: Figures are grams of protein per 100g of minced beef. Proportions are based on the number of observed purchases of all minced beef products in 2006 in the TNS data. The line represents the aggregate EFS intensity of 20.3g

**Figure 4: Unsaturated fat per 100g in peanut butter products, weighted by number of purchases**
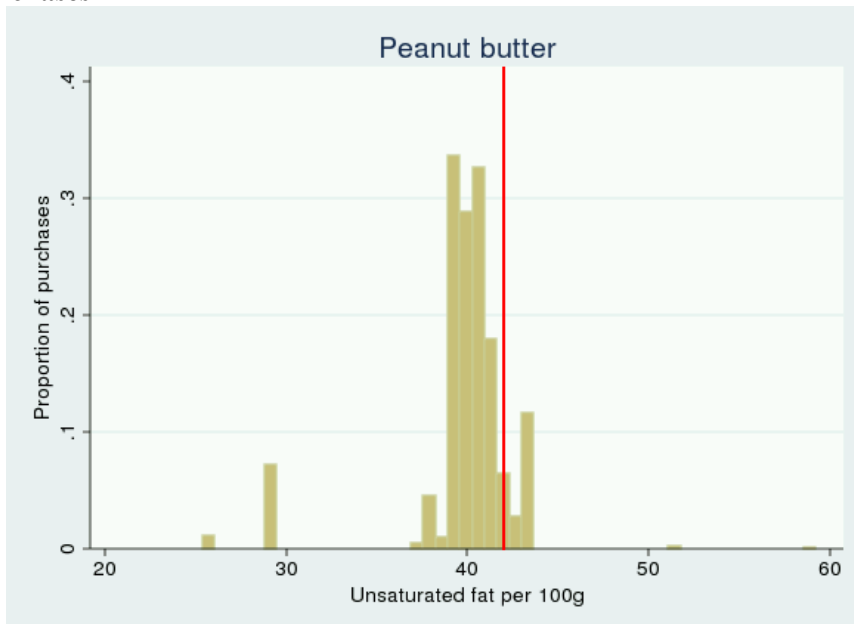
Notes: Figures are grams of unsaturated fat per 100g of peanut butter. The proportions are based on the number of observed purchases of all peanut butter products in 2006 in the TNS data. The line represents the aggregate EFS intensity of 42.0g.

Figure 5: Distribution of proportion of energy from carbohydrates across households, measured using different duration of sample period
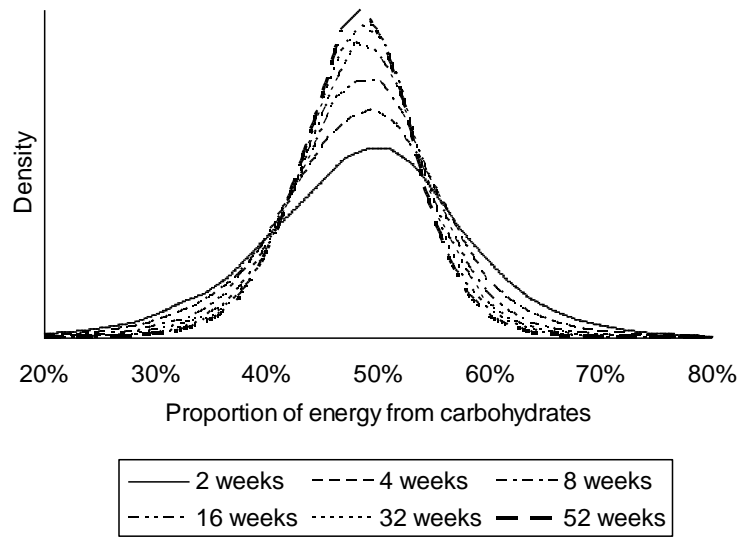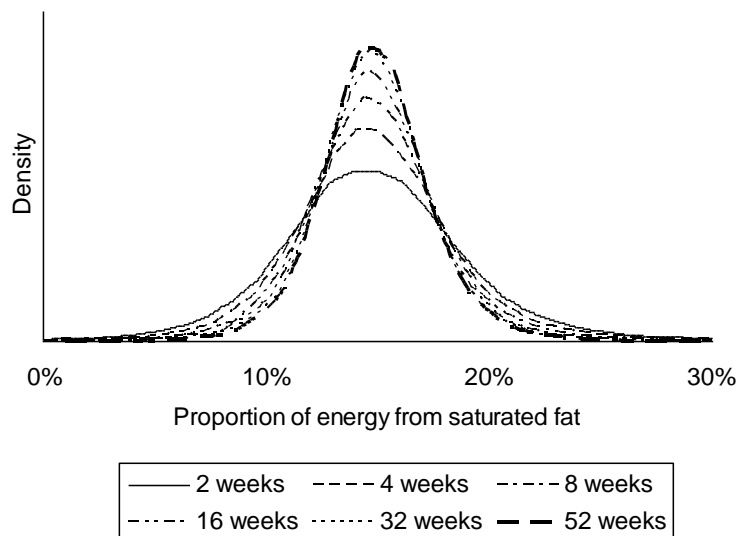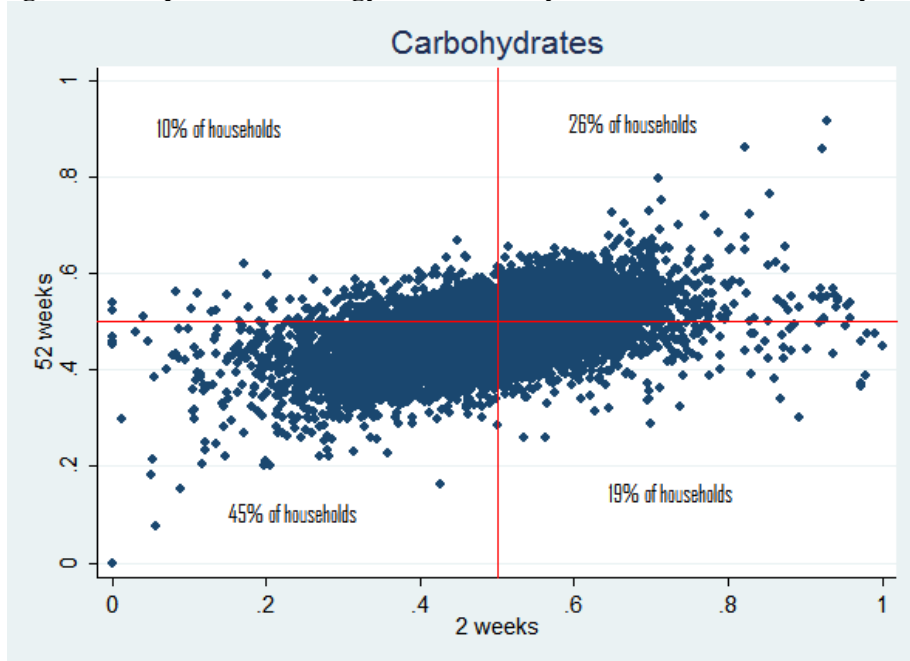

Notes: For each duration except fifty-two weeks, we randomly select a period for each of the 15,282 households that remained in the TNS data throughout 2006.

Figure 6: Distribution of proportion of energy from protein across households, measured using different duration of sample period
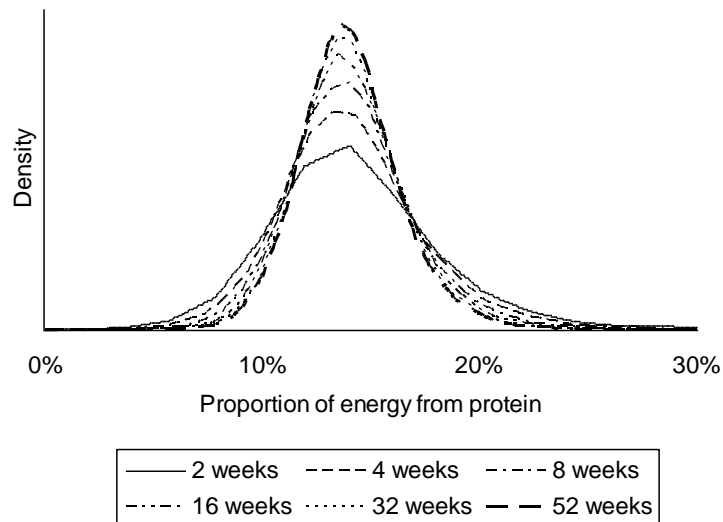
Notes: For each duration except fifty-two weeks, we randomly select a period for each of the 15,282 households that remained in the TNS data throughout 2006.

Figure 7: Distribution of proportion of energy from unsaturated fat across households, measured using different duration of sample period
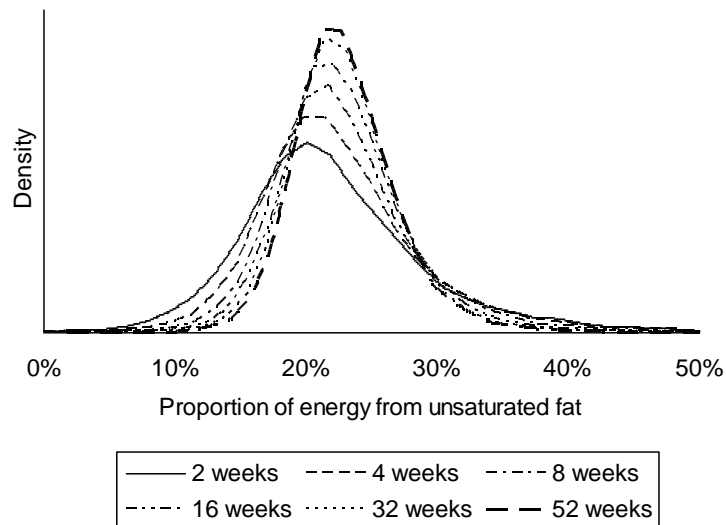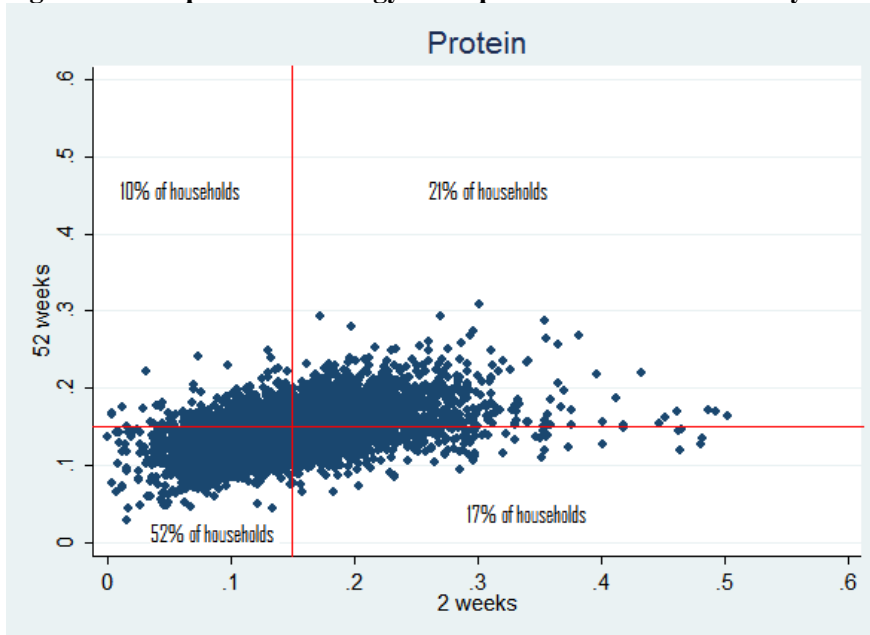

Notes: For each duration except fifty-two weeks, we randomly select a period for each of the 15,282 households that remained in the TNS data throughout 2006.

Figure 8: Distribution of proportion of energy from saturated fat across households, measured using different duration of sample period

Notes: For each duration except fifty-two weeks, we randomly select a period for each of the 15,282 households that remained in the TNS data throughout 2006.

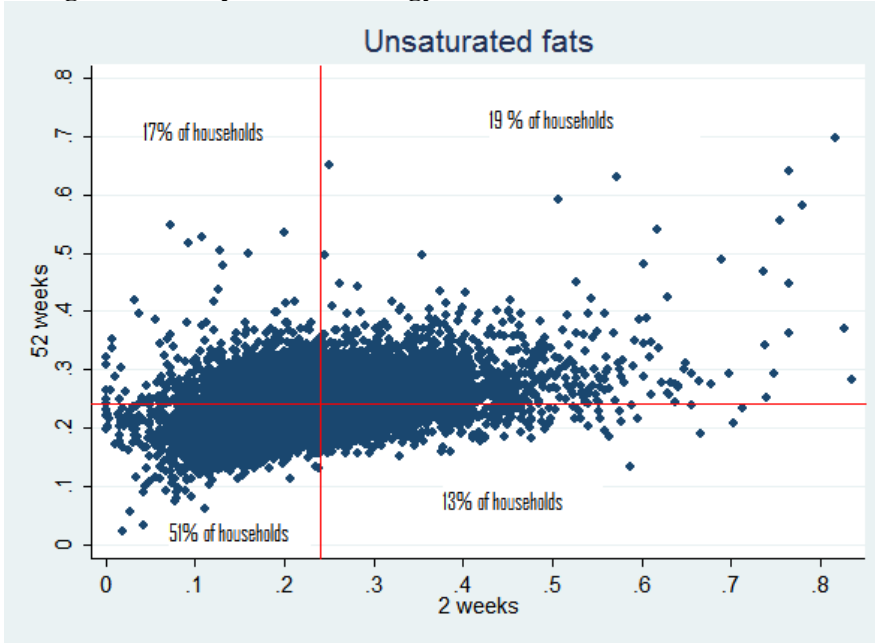**Figure 9: Proportion of energy from carbohydrates relative to Dietary Reference Value**

Notes: The horizontal and vertical lines correspond to Department of Health Dietary Reference Values. The percentage of households in each quadrant is reported. For the 2-week duration we randomly select a period for each of the 15,282 households that remained in the TNS data throughout 2006.

Figure 10: Proportion of energy from protein relative to Dietary Reference Value
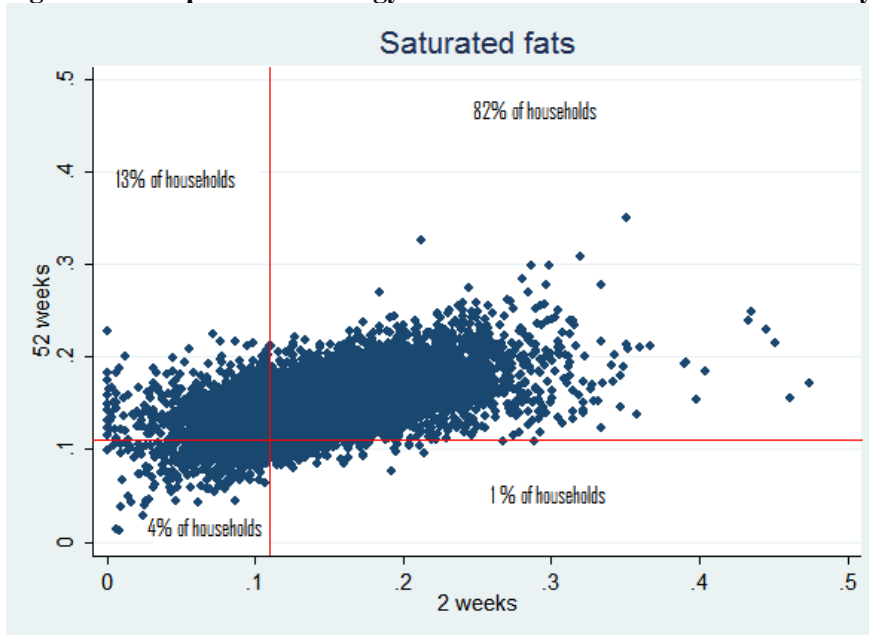

Notes: The horizontal and vertical lines correspond to Department of Health Dietary Reference Values. The percentage of households in each quadrant is reported. The figure excludes three households that buy over 60% of their energy as protein using the two week duration. For the 2-week duration we randomly select a period for each of the 15,282 households that remained in the TNS data throughout 2006.

8

Figure 11: Proportion of energy from unsaturated fat relative to Dietary Reference Value

Notes: The horizontal and vertical lines correspond to Department of Health Dietary Reference Values. The percentage of households in each quadrant is reported. The figure excludes three households that buy over 80% of their energy as unsaturated fat using the two week duration. For the 2-week duration we randomly select a period for each of the 15,282 households that remained in the TNS data throughout 2006.

Figure 12: Proportion of energy from saturated fat relative to Dietary Reference Value

Notes: The horizontal and vertical lines correspond to Department of Health Dietary Reference Values. The percentage of households in each quadrant is reported. The figure excludes one household that buys over 50% of its energy as saturated fat using the two week duration. For the 2-week duration we randomly select a period for each of the 15,282 households that remained in the TNS data throughout 2006.