Exploring the role of consumer data for food in national survey reporting

Morris MA[*1], Clarke GP[†1] and Birkin MH[‡1]

[1]School of Geography, University of Leeds

January 9, 2014

Summary

National data collected to aid understanding of spending patterns associated with food consumption and nutrition in the UK are reported in the Family Food module of the Living Costs and Food Survey. This survey data is used to better understand our society, including the geographies of food spending and informs the Consumer Price Indices. Such surveys require a nationally representative sample of volunteers to report their food spending and consumption patterns. This paper explores the role in which big data relating to food purchases could supplement such surveys and how reports using consumer data compare to survey data geographically.

KEYWORDS:

Consumer data
National survey
Living Costs and Food
Food Consumption
Nutrition

1. Extended Abstract

1.1 Background

Collecting data from households using survey methods is a costly and time consuming process. The Office for National Statistics (ONS) carry out this process for the Living Costs and Food Survey (LCFS) on an annual basis, recruiting around 6000 households in the UK (around 150 of these being in Northern Ireland) (Office for National Statistics 2012). The European Standard Classification of Individual Consumption by Purpose (COICOP) (Unitied Nation Statistics Division 2015) is used within the LCFS to categorise spending. In conjunction with DEFRA the Family Food module of the survey is produced, focussing on household spending on food and drink (Department for Environment Food and Rural Affairs 2012). The nutrient content of family diets is derived from the food spending records.

Food price inflation has been higher than general inflation since 2007 meaning that spending on food is a higher proportion of household budgets than previously (Office for National Statistics 2012). During the recent recessionary times, cost savings have been made in food purchases (United States Department of Agriculture 2011, Crossley, Low et al. 2012). Diet is the leading modifiable risk factor in non-communicable chronic diseases such as Obesity, Type 2 diabetes, Cardiovascular disease and some cancers (Institute for Health Metrics and Evaluation 2013, US Burden of Disease Collaborators

[*] m.morris@leeds.ac.uk
[†] g.p.clarke@leeds.ac.uk
[‡] m.h.birkin@leeds.ac.uk

2013). It is widely published that a healthy diet is more expensive than a less healthy one (Rehm, Monsivais et al. 2011, Morris, Hulme et al. 2014) so it is wholly possible that cost savings in diet, at the expense of diet quality, could impact health in the longer term. Consideration of the opportunity cost of making such choices is unlikely to be considered by the consumer, but reports such as the Family Food Survey allow the government and researchers to monitor consumer behaviour in relation to food.

Consumer transaction data from UK retailers for food and drink purchases are collected from millions of households, often with the ability to link to loyalty card information containing demographic characteristics (Felgate, Fearne et al. 2012). The Consumer Data Research Centre (CDRC), an ESRC data investment, has been established to broker such data between retailers and researchers in order to tackle global challenges over the coming years (Consumer Data Research Centre 2015). Better understanding of consumer behaviour in respect to food purchases could contribute to improved public and preventative health.

Geographies of consumption are typically reported at a large geographical unit such as Government Office Region in the UK. This ensures anonymity but aggregates to such a scale that wide generalisation occurs and pockets of certain dietary behaviours are lost. Some surveys present results for geodemographic group which allows specific groups to be more easily pinpointed spatially, but the generalisation occurs according to demographics through segmentation of the whole country into a defined number of groups. Using Big Data which includes supermarket loyalty card records results at a small spatial scale can be generated, but care regarding the level which these are reported needs to be taken to ensure anonymity is maintained.

Using consumer data as a measure of diet is not a new concept. Much research has been carried out using a 'basket analysis' approach whereby a typical basket of food is defined, often with the components required to meet the national dietary guidelines (Mooney 1990, Sooman, Macintyre et al. 1993, Larsen and Gilliland 2009, Cummins, Smith et al. 2010, Drewnowski, Aggarwal et al. 2012). Important concerns relating to the cost of the basket, and hence the cost of meeting dietary guidelines can be explored. The concept of Big Data provides an opportunity to build on existing research on an amplified scale.

1.2 Methods

While consumer data is 'Big Data' and presents many exciting research opportunities, there are also challenges to overcome. Building relationships with retailers is essential. It is likely that data will be available only from certain retailers, especially in the beginning. This data will likely over-represent certain socioeconomics groups within a population. The data will then require weighting such that it becomes representative. While the numbers will be much higher than those in national survey samples, with wide geographic reach, they may not be evenly spatially distributed and again this will need to be accounted for in any reporting and analysis.

This paper will discuss the methods employed for manipulation of consumer data into shape for research purposes as a first step for researchers; a process which will be contributed to from previous lessons learned and continue to be refined over time. Nutrient costs will be presented in the format of the Family Food Survey. Utilising the same classification will allow for comparison of the reports generated from different data sources (survey verses consumer data). It will be possible to perform post hoc power calculations for the ability to detect nutrient differences which are clinically important for health for each data source.

Research into the effect of nutrient consumption and health outcomes typically uses dietary data from self-reported food records, be it a 4 day weighed food diary, a series of 24 hour dietary recalls or a Food Frequency Questionnaire which aims to record typical diet over defined period of time. Dietary self-reporting is subject to a range of bias, such as the under-reporting by those who are overweight or

obese. Using consumer data provides real data on food which has been purchased, but without dietary consumption records. Assumptions need to be made regarding the distribution of food consumption within a household and issues such as how much of food is wasted, or remains in the store cupboard until another week or month. Methods to account for these will be discussed.

Geographies of consumption will be explored and results compared to those of the Family Food Survey. If possible inter retailer differences will be explored spatially, according to Government Office Region and by Geodemographic category. The retailer generated report will be cross referenced to a synthetic population generated by spatial microsimulation techniques using small area microdata from the 2011 census. This nationally representations synthetic population will enable us to better understand the bias resulting from regional variation in retailer distribution, shown in figure 1 and also sociodemographic variations in customers to these stores, shown by education level in figure 2. Spatial interaction modelling methods will be used to explain such variations.
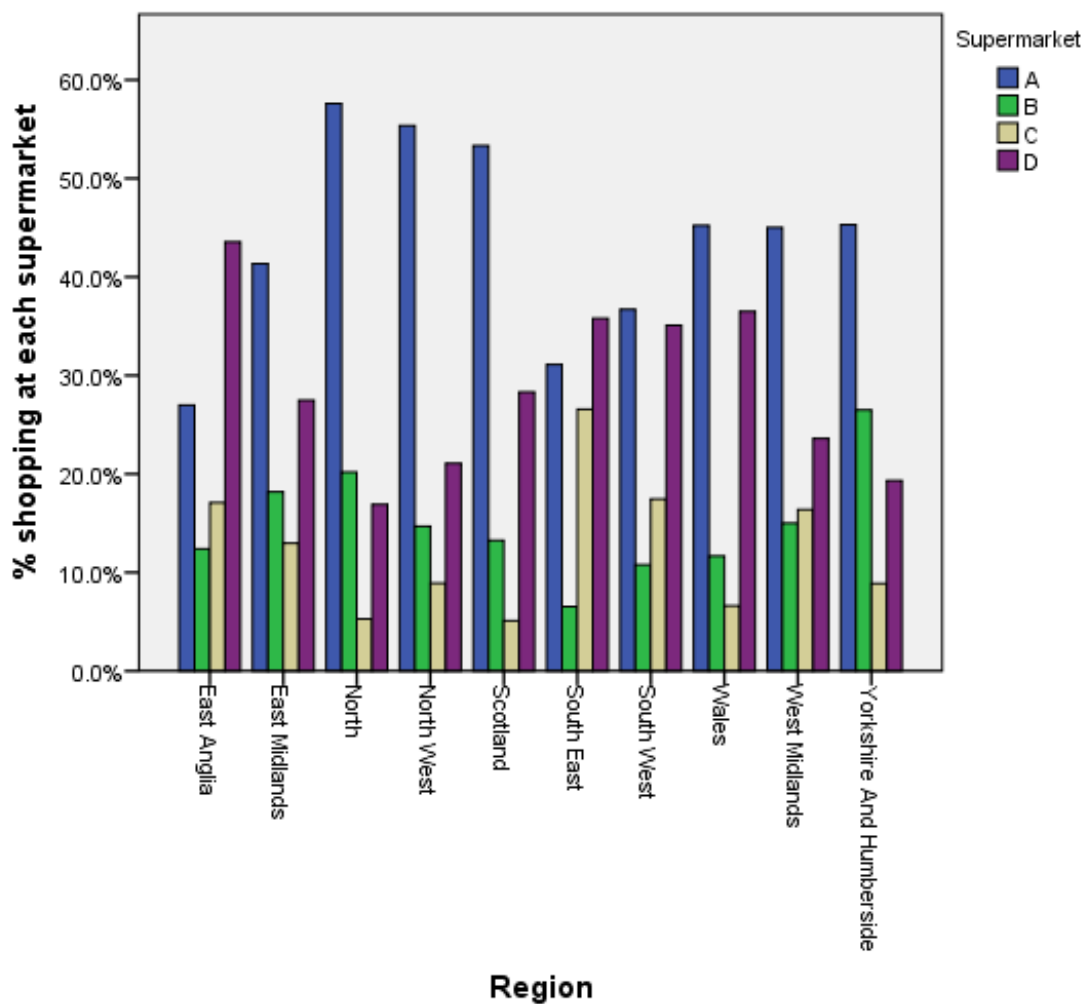

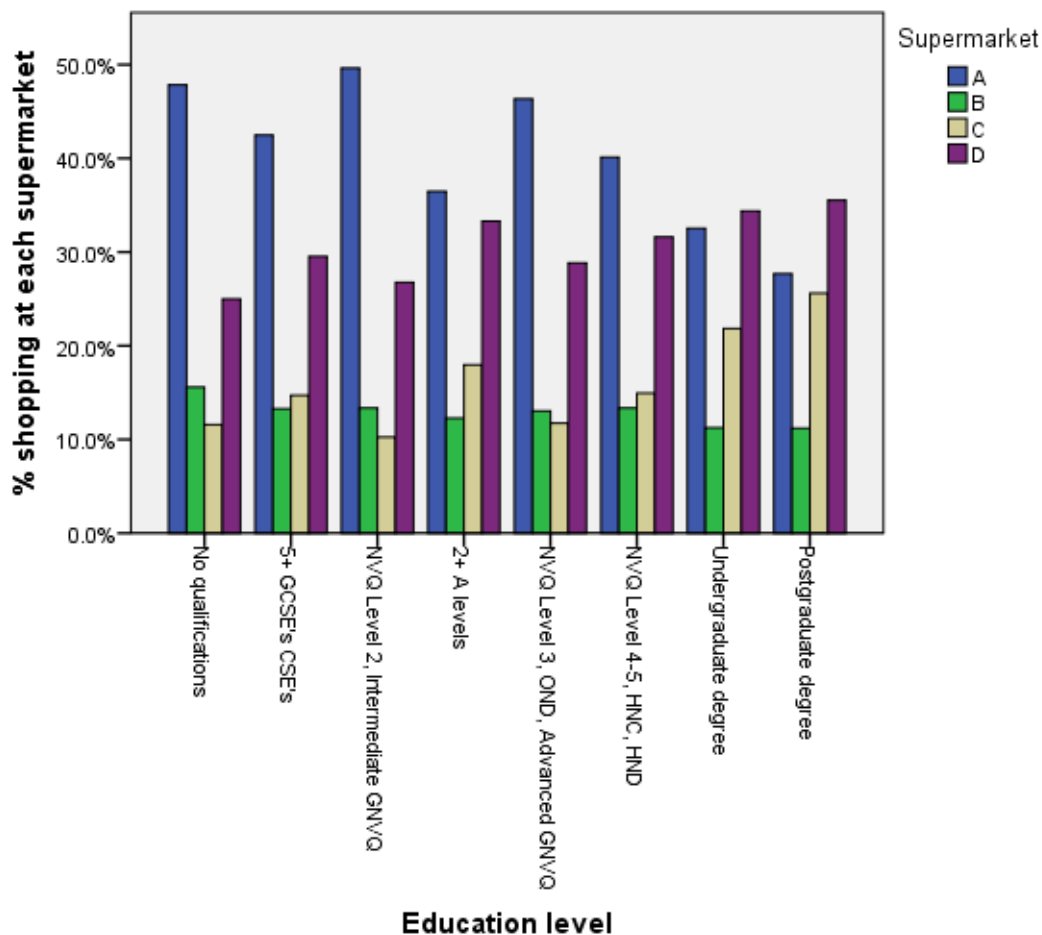Figure 1 – Use of different supermarkets by region, derived from Acxiom survey data 2007.

Figure 2 – Percentage of shoppers at different supermarkets by educational level, derived from Acxiom survey data 2007.

1.3 Implications

Results could open up new channels of thought regarding how to report and analyse behaviour such and food purchasing and consumption. It is possible that results from Big Data could provide greater reliability in survey estimates from increased sample size power and a decrease in reporting bias, in a more timely manner, as retail data is collected daily, without a survey data collection time lag. It could provide policy makers with greater opportunity and confidence in implementing new initiatives to influence diet and subsequent health. Better understanding of the geographies of spending and consumption of food are important to policy makers, health professionals and retailers.

1.4 Future work

Real time experiments could be carried out in conjunction with retailers to pilot important interventions such as: the impact of taxing certain foods; changing product prices; altering availability of products; simulated effects of change in quality; impact of guidance from national campaigns; or the effect of manipulating local social norms. This provides an exciting alternative to virtual supermarket experiments and advanced modelling techniques.

Acknowledgements

Biographies

Michelle Morris
Michelle is a postdoctoral research fellow in the Consumer Data Research Centre at the University of Leeds. Her primary research interests are in spatial variations in diet and health. Michelle is an interdisciplinary researcher with a background spanning, spatial analysis and policy, nutritional epidemiology and health economics.

Graham Clarke
Graham is professor of business geography at the Centre for Spatial Analysis and Policy at the University of Leeds. Graham's research interests include GIS, urban services, retail and business geography, geographies of health, urban modelling and continued professional education.

Mark Birkin

Mark is Director of the Consumer Data Research Centre, based at the University of Leeds where he is Professor of Spatial Analysis. His major research interests are in simulating social and demographic change within cities and regions and in understanding the impact of these changes on the need for services.

References

Consumer Data Research Centre. (2015). "Consumer Data Research Centre." from http://cdrc.ac.uk/.

Crossley, T., H. Low and C. O'Dea (2012). Household Consumption Through Recent Recessions. Institute for fiscal studies - working paper W11/18.

Cummins, S., D. M. Smith, Z. Aitken, J. Dawson, D. Marshall, L. Sparks and A. S. Anderson (2010). "Neighbourhood deprivation and the price and availability of fruit and vegetables in Scotland." Journal of Human Nutrition & Dietetics **23**(5): 494-501.

Department for Environment Food and Rural Affairs (2012). Family Food 2012. D. f. E. F. a. R. Affairs.

Drewnowski, A., A. Aggarwal, P. M. Hurvitz, P. Monsivais and A. V. Moudon (2012). "Obesity and supermarket access: proximity or price?" Am J Public Health **102**(8): e74-80.

Felgate, M., A. Fearne, S. DiFalco and M. G. Martinez (2012). "Using supermarket loyalty card data to analyse the impact of promotions." International Journal of Market Research **54**(2): 221-240.

Institute for Health Metrics and Evaluation (2013). The Global Burden of Disease: Generating Evidence, Guiding Policy – European Union and Free Trade Association Regional Edition.

Larsen, K. and J. Gilliland (2009). "A farmers' market in a food desert: Evaluating impacts on the price and availability of healthy food." Health & Place **15**(4): 1158-1162.

Mooney, C. (1990). "Cost and availability of healthy food choices in a London health district." Journal of Human Nutrition and Dietetics **3**: 111-120.

Morris, M. A., C. Hulme, G. P. Clarke, K. L. Edwards and J. E. Cade (2014). "What is the cost of a healthy diet? Using diet data from the UK Women's Cohort Study." J Epidemiol Community Health

68(11): 1043-1049.

Office for National Statistics. (2012). "Living Costs and Food Survey, 2012." <u>Living Costs and Food Survey (Expenditure and Food Survey)</u> Retrieved January, 2015, from <u>http://discover.ukdataservice.ac.uk/catalogue/?sn=7472&type=Data%20catalogue</u>.

Rehm, C. D., P. Monsivais and A. Drewnowski (2011). "The quality and monetary value of diets consumed by adults in the United States." <u>American Journal of Clinical Nutrition</u> **94**(5): 1333-1339.

Sooman, A., S. Macintyre and A. Anderson (1993). "Scotland's health--a more difficult challenge for some? The price and availability of healthy foods in socially contrasting localities in the west of Scotland." <u>Health Bull (Edinb)</u> **51**(5): 276-284.

United States Department of Agriculture. (2011). "Food Spending Adjustments During Recessionary Times." from <u>http://www.ers.usda.gov/amber-waves/2011-september/food-spending.aspx</u>.

Unitied Nation Statistics Division. (2015). "COICOP (Classification of Individual Consumption According to Purpose)." Retrieved January, 2015, from <u>http://unstats.un.org/unsd/cr/registry/regcst.asp?Cl=5</u>.

US Burden of Disease Collaborators (2013). "The state of US health, 1990-2010: burden of diseases, injuries, and risk factors." <u>The Journal of the American Medical Association</u> **310**(6): 591-608.