**The use of scanner technology for collecting expenditure data in mixed-mode social science surveys**

## Aim

Traditionally, micro-level household expenditure data has been gathered through spending diaries or through 'recall' questions, asking people how much they spent on certain items over the previous few months. Both have problems. Diaries can be comprehensive but will cover only a limited period due to the time-intensive nature of the data collection process, which can lead to deliberate or accidental omissions in the data and problems with infrequency of purchase for many durable and semi-durable items. Recall questions are quicker to answer but may contain errors as people forget their expenditures.

An alternative is to use a barcode scanner in the home to record information on products purchased quickly and accurately. Such data is available from commercial sources, with the added advantage of being longitudinal rather than cross-section. However, these data have not been used for academic research in the UK, but could in principle enable research to be carried out on a broad range of important and policy relevant topics.

The aim of this research is to compare scanner data from market research firm Taylor Nelson Sofres (TNS) covering household grocery purchases to more traditional data sets to assess the usefulness of this mode of spending data collection. We aim to explore survey measurement issues with longitudinal spending data such as response, participant fatigue and attrition, repeated collection of demographic information and so on. We wish to fully understand any issues which arise and their implication for work that could be done with this data in the future, and for other data sets that wish to incorporate similar technologies into their design. It will also open up the possibility that the wider research community will begin to use similar data, both by raising awareness of its existence and by providing a detailed account of its strengths and weaknesses.

## Work so far

The project started in March 2008. Early work has focused on preparing the TNS data for analysis: we have information spanning six years, with an average week of data containing 15,000 – 25,000 households, 50,000 different products and 750,000 purchases. To compare the expenditure data requires us to map these products to individual expenditure codes in the UK Expenditure and Food Survey. We have also needed to clean up the expenditure and demographic information in order to be able to make comparisons across time and across surveys. Given the scale of the data, this has been a time-consuming process.

Having completed the data preparation phase, we are now starting our comparisons of the distributions of expenditures and demographics in the scanner data and traditional data sets. We have begun by comparing expenditures at a very detailed level though it is likely we will need to consider more aggregate categories to make the analysis tractable. We have also started to examine how expenditures vary over time within households to see whether there is evidence of 'panel fatigue' and the impact of attrition.