Enhancing the collection and analysis of survey and administrative data by linking

Smith, P\textsuperscript{1*}, Moore, J\textsuperscript{1}, and Durrant, G\textsuperscript{1}

\textsuperscript{1}University of Southampton

**Background**

Linking survey and administrative data can enhance their collection and analysis. By linking the issued sample to administrative data, the representativeness of a survey can be assessed during and at the end of data collection. An understanding of the (lack of) representativeness can inform adaptive and responsive survey designs to improve representativeness and/or reduce survey costs. It can also aid the assessment of potential survey non-response bias and the implementation of strategies to adjust for such bias during analysis. Comparison of responses to the same question in two linked datasets can help identify problematic questions. This information can then be used to improve question wording to reduce measurement error in future surveys and to account for such error during analysis.

**Objectives**

To assess how the representativeness of a survey changes after each attempt to contact the interviewees, and to assess the nature and extent of differential reporting between a face-to-face survey and the self-completion 2011 UK Census.

**Methods (including data)**

We present our recent research developing representativeness indicators based on the coefficient of variation of the response propensities calculated after each attempt to contact the interviewees and illustrate their use using Labour Force Survey (LFS) data linked to contemporaneous data from the 2011 Census. We also describe our work on differential reporting between the LFS and the Census, focusing on the reporting of highest qualification and ethnicity.

**Findings and Conclusions**

Our results suggest that considerable savings can be made by reducing the number of attempts to obtain a response without a detrimental effect on representativeness and that face-to-face interviews obtain more accurate information than self-completion questionnaires.