

**Title:** *Missing Income Data in the Millennium Cohort Study: Evidence from the First Two Sweeps*

**Author(s):** Denise Hawkes and Ian Plewis

**Institution/Country:** Centre for Longitudinal Studies (CLS) and Institute of Education, University of London

The Millennium Cohort Study (MCS) is the fourth in a series of internationally renowned cohort studies in the UK. It includes 18,818 babies in 18,552 families born over a 12 month period and living in selected UK wards at age 9 months. Areas with high proportions of Black and Asian families, disadvantaged areas and the three smaller UK countries are all over-represented in the sample which is disproportionately stratified and clustered. The first and second sweeps took place when the cohort members were 9 months and 3 years old. In addition, at sweep two, families who were living in sampled areas with a child of the appropriate age but who were not located at the first sweep were introduced. These “new families” tend to be more mobile than those already part of the MCS. Partners were interviewed whenever possible and detailed questions about individual and household income were included in both sweeps.

There are four ways in which income data can be missing. There was unit non-response at sweep one such that the response rate then was 72%. There was further partner non-response; the partner response rate at sweep one, among respondent families with partners, was 88%. In addition there was item non-response for income: about 6% of main respondents and 6% of partners did not provide income data at sweep one. Moreover, there was attrition between sweeps one and two: 79% of eligible cases responded at sweep two. The correlates of unit and partner non-response at sweep one are set out in Plewis (2004); the evolution of the sample from sweep one to sweep two is described in Plewis and Ketende (2006).

The paper will address the following questions:

- (i) Are there (a) within household and (b) within individual correlations for missing income data?
- (ii) Is a female interviewer more successful than a male interviewer in getting responses to income questions from main respondents and their partners?
- (iii) Is there a systematic tendency for income data to be missing at sweeps one and two over and above what we know about unit and partner non-response?
- (iv) Is attrition at sweep two related to (a) household income at sweep one; (b) the failure to provide income data at sweep one?

The paper will conclude by considering the implications for statistical modelling and future data collection of our findings on the patterns and correlates of income non-response at both sweeps.

Plewis, I. (ed.) (2004) Millennium Cohort Study First Survey: Technical Report on Sampling (3rd. Edition). Centre for Longitudinal Studies: London.