Adjusting for Within Household Dependence

Owen Abbott (ONS) and James Brown (UoS)

1) Introduction

The work that led to the DSE bias adjustments (called the dependence adjustment) in the 2001 ONC assumed that within household dependence would be minimal because the within household coverage of households found by the Census Coverage Survey (CCS) would be very high (see Brown *et al*, 2006). However, this is a strong assumption and as we prepare for 2011 it is an assumption we would like to be able to test. The approach to dependence adjustment developed and tested for 2011 can reflect within household dependence within the framework but so far the within household odds ratio has always been set to one. In this paper we explore an approach to estimating that odds ratio so that the overall dependence adjustment can be set to reflect any within household dependence.

2) Source of Information

As with previous Censuses, a linkage between households sampled in the main ONS surveys at the time of the 2011 Census and the corresponding census returns is envisaged to study the characteristics of survey non-responders. However, where there is a survey response, we would expect the interviewer to achieve a *near perfect* listing of the household members. Therefore, the linkage with the 2011 Census returns for the households gives us a measure of the census coverage within households[1] that is completely independent of the CCS. This would likely be possible for high levels of geographic aggregation by hard-to-count by broad age-sex groups as was done for the dependence adjustment in 2001 (see Brown *et al*, 2006). The matching required for this analysis is planned to coincide with the census processing timetable, allowing an assessment of and potential adjustment for within household dependence within the census production timeframe.

3) Using the Information

Based on the linkage to other social surveys, we estimate the within household coverage of the census as $\hat{C}_{SS}$. After matching between the census and the CCS (and collapsing over the same groupings as for $\hat{C}_{SS}$) we observe the following information for individuals within **matched** households

		CCS		
		Counted	Missed	
Census	Counted	n_{11}	n_{10}	n_{1+}
	Missed	n_{01}		
		n_{+1}		

[1] There are difficulties with definitions of household membership, but these need to be addressed to allow for the non-response study so should not be considered insurmountable.

and can estimate the within household coverage of the census as $\hat{C}_{CCS} = \frac{n_{11}}{n_{+1}}$. (Given that we will be collapsing over geographic areas with differing historical coverage the CCS sampling rates are very different and we will need to account for this, either by allowing for the effects or by including the CCS weights in creating the counts in the table.) Under independence the count n_{10} does not need to be zero as $\hat{C}_{CCS} = \frac{n_{11}}{n_{+1}}$ is estimating the true within household coverage $C_{CCS} = \frac{n_{1+}}{n_{++}}$. However, a non-zero value exposes us to the possibility of an impact from within household dependence. We can now compare $\hat{C}_{SS}$ with $\hat{C}_{CCS}$ but of course there will be variability in both. Each one is an independent estimate of the underlying census coverage, which is just a proportion, and as an approximation we can get a standard error on each assuming simple random sampling. Given we detect a difference, this implies within household dependence between the census and the CCS. Therefore, we can use $\hat{C}_{SS}$ to construct an estimate of n_{++} as $\hat{n}_{++} = \frac{n_{1+}}{\hat{C}_{SS}}$ and therefore an estimate of the count n_{00}, those missed by both the census and the CCS, is given by

$$\hat{n}_{00} = \frac{n_{1+}}{\hat{C}_{SS}} - n_{11} - n_{10} - n_{01}. \tag{3}$$

Therefore, as the odds ratio for a two by two table is just $\gamma = \frac{n_{11}n_{00}}{n_{10}n_{01}}$, we can now estimate the odds ratio due to within household dependence as

$$\hat{\gamma} = \frac{n_{11}\left(\frac{n_{1+}}{\hat{C}_{SS}} - n_{11} - n_{10} - n_{01}\right)}{n_{10}n_{01}}. \tag{4}$$

The odds ratio estimate given by (4) can then be embedded in the standard dependence adjustment we have already tested. This will potentially make the adjustment even more age-sex specific within the broad groups we define for the estimation and comparison of $\hat{C}_{SS}$ and $\hat{C}_{CCS}$.

4) Issues to Consider

1. Do we need individual level matching or can we achieve an estimate of $\hat{C}_{SS}$ by simply comparing distributions of household size? Are there administrative sources that might create an alternative? These might provide other (or earlier) evidence to confirm the existence/extent of within household dependence.
2. Collapsing over areas in the CCS will always tend to create the appearance of dependence due to heterogeneity, which is not in the actual estimate. However, we are

arguing that within household coverage of the Census will vary less across geography, once we control for broad age-sex groups and hard-to-count.

3. Should this be a one-sided adjustment? It is likely that any alternative source will have some coverage issues (i.e. less than perfect) and there will be definitional issues. Therefore, if we are careful with definitions and exclude any individuals we are unsure of, our estimate $\hat{C}_{SS}$ will be 'conservative' so it makes sense that any adjustment should only go in one direction.