UK Data Service

Answers: De-identification of quantitative data

The table below shows some assessed variables (variables commonly assessed for disclosure risk such as age, community, but also variables for which local knowledge is essential to indicate risk) with identified disclosure risk and recommendations for actions.

Variables	Disclosure risk	Action
Community	Low frequency counts for all named communities, respondents who gave answers very easily identifiable (especially in combination with other variables).	Exclude variable from dataset
Age	Low counts of older respondents over 75 years old	Top-code age >= 75 as '75 and over'
Main occupation during last 12 months	Low counts of very specific occupations.	Occupations aggregated into standard occupation codes
Ethnicity of the Household Head	Low counts of specific ethnicities.	Recode the low-frequency responses (all responses but 'Mamprusi' and 'Builsa') into 'Other'.
Household's primary type or energy/fuel used for cooking	Very low counts for 'Gas/LPG' and 'Electricity-solar panel' responses may lead to household identification (especially if combined with other datasets)	Recode all responses into the following main categories: 1 - 'Firewood'; 2 - 'Electricity-based'; 3 - 'Charcoal'; 4 - 'Other', 5 - 'Don't know'; 6 - 'NA/missing'.
Main material of the wall of the house	A number of low-frequency responses; exterior features (households/buildings easily identifiable).	As the main material of the wall refers to the exterior of a building, it may be advisable to recode the low-frequency and 'Other' variables into 'Other (incl. wood-based and stone-based') and retain the remaining groups
Crops grown on plots	A number of low-frequency specific responses for each variable.	Variables are recoded into crop categories