

## 7.8 Item Response

*Adriaan Kalwij and Arthur van Soest*

As common in household surveys, respondents sometimes answered questions with “I don’t know” (DK) or “I’d rather not say” (RF, refusal). This behaviour is called item non-response. For an overwhelmingly large majority of the variables in SHARE, item non-response is a minor problem since the percentage of DKs or RFs is quite small. For example, there is hardly any item non-response in physical or mental health variables, in well-being, labour force status and job satisfaction, or in basic demographics and information on children. Somewhat larger item non-response rates are found for qualitative questions on pension entitlements, expectations, asset ownership or the nature of the assets.

The type of questions that suffers substantially from item non-response are questions on amounts of income, expenditure, or values of assets. In this respect, however, SHARE does not differ much from comparable surveys like ELSA or HRS. For example, owners of shares of stock or stock mutual funds are asked the total value of their (household’s) shares of stock and stock mutual funds. In SHARE, 30.2% of the owners answer DK or RF, compared to 35.0% in HRS wave 2002.

Respondents answering DK or RF are asked a number of subsequent questions on whether the amount is larger than, smaller than, or about equal to a given amount. This so-called unfolding bracket design was already used in HRS 1992 and proved to be an effective way to collect categorical information on the initial non-respondents. For example, with bracket questions on the amounts €25,000, €50,000 and €100,000, for those who go through all the bracket questions, we know whether the amount is less than €25,000, about €25,000, between €25,000 and €50,000, about €50,000, etc. Like in HRS, a large fraction of initial non-respondents appear to be willing to answer the bracket questions. For examples, for shares of stocks and stock mutual funds, 45.4% of initial non-respondents in SHARE complete the brackets, compared to 41.2% in HRS. For 16.5% of all owners in SHARE, there is no information on the amount at all, compared to 18.6% in HRS. Thus SHARE compares favourably to HRS in this respect, something that is generally also found for other amount questions.

For studies that use income or income components, wealth or wealth components, etc., as one of the right-hand variables, missing information on one of these variables is a problem. Deleting observations with missing information is often an unattractive option for two reasons. The first reason is that a smaller sample size results in an efficiency loss. The second reason is that deleting missing data may yield biased inferences when item non-response is related to the variable of interest. For instance, the reason for item non-response may be related to the same factors that drive income or health of the respondent and deleting missing data would then lead to a selective sample.

Therefore, instead of deleting missing data, the missing values are replaced by imputed values, i.e., observed values of other respondents that are similar to the respondent considered in certain relevant aspects. Many imputation methods exist. For the data release used by all papers in this volume, we followed the procedure of Hoynes et al. (1998). Imputations were first done recursively for a small set of core variables (income from employment, self employment or public pensions, value of owner-occupied housing, amount held in shares of stock and stock mutual funds, amount held in checking and saving accounts, and food consumption). This is done to guarantee that imputations respect the correlation structure of these variables. For example, respondents with missing food consumption but with high (observed or imputed) earnings, were assigned an observed (probably relatively

high) food consumption amount of another respondent with similarly high earnings but with observed food consumption. The imputed values are flagged (i.e., an indicator variable is constructed indicating the level of imputation) and flags and imputed variables will be included in the public release of the data. More refined imputation methods will be applied to later data releases.

## 7.9 Computing a Comparable Health Index

*Hendrik Jürges*

Subjective data, such as self-assessed health can be subject to cross-country bias for several reasons. However, there is a fairly straightforward possibility to compute a single measure of health, that is comparable across countries. The main requisite are objective data on the respondents health: self reported diagnosed chronic conditions, mental illnesses, symptoms (especially pain), or functional limitations. If available, one also uses medical records, and measurements and tests like blood samples, grip strength, balance, gait speed, etc. The absence of any conditions, symptoms, or limitations, implies perfect health, i.e. an index value of 1. The presence of a condition reduces the health index by some given amount or %age, the so-called disability weight. The disability weight of each condition or symptom is assumed to be the same for each respondent.

Disability weights are often derived by expert judgements or surveys specialised to elicit health preferences, using time trade-off or standard gambles. In SHARE, we are able to compute disability weights from within our sample (Cutler and Richardson 1997) by estimating ordered probability (e.g. probit) models of self-reported health (which ranges e.g. from „excellent“ to „poor“) on a large number of variables representing chronic conditions, symptoms, ADL problems, depression, physical functioning, height, weight, and cognitive functioning. We can also include our measures of grip strength and walking speed, and basic demographic variables like age and sex. The health index is then computed as the linear prediction from this regression (the latent variable), normalised to 0 for the worst observed health state (often referred to as „near death“) and 1 for the best observed health state (referred to as „perfect health“). This procedure implies disability weights for each condition or impairment that are equal to the respective (also normalised) regression parameters. Since the variable on which we base this measurement is self-reported health itself (and thus potentially subject to cross-cultural bias), we account for country specific reporting styles by modelling the latent variable thresholds as a function of country of residence (i.e. we basically have fixed country effects at each threshold). Thus thresholds are allowed to vary across countries, while disability weights are constrained to be the same in each country.

## 7.10 Income Imputation

*Omar Paccagnella and Guglielmo Weber*

*The Definition of Income:* Total income is the sum of some incomes at the individual level and some at the household level. The basic definition used in the SHARE project reflects money income before taxes on a yearly base (2003) and includes only regular payments. Lump-sum payments and financial support provided by parents, relatives or other people are not included.

The available data at the individual level include: income from employment; income from self-employment or work for a family business; income from (public or private) pensions or invalidity or unemployment benefits; income from alimony or other private regu-