Cross-nationally Equivalent Data Accuracy?

Item nonresponse analysis of the Generations and Gender Surveys and implications for international comparative research

Andrej Kveder[1], Nicole Hiekel[1] and Julia Mikolai[2]

[1] *Netherlands Interdisciplinary Demographic Institute* [2] *Utrecht University*

Is data accuracy equivalent across countries and cultures? Do differential patterns of data quality have an impact on the results of international comparative research? This paper intends to analyse item nonresponse patterns and its determinants across the Generations and Gender Surveys (GGS) from eight countries. A special emphasis will be placed on the cross-national equivalence comparison of the item nonresponse patterns and its potential implications for international comparative research.

One of the main postulates of comparative research is that the differences in findings observed across countries can actually be attributed to heterogeneity in country-level characteristics. However, recent research in survey measurement quality and questionnaire design has shown that the observed differences in main variables of interest can also result from failed functional equivalence of measures or from a differential quality of measurement (e.g. Saris and Gallhofer, 2007). Item nonresponse errors and measurement errors are closely related or even interrelated terms describing aspects of total survey error. The emphasis of this paper is on item nonresponse and thus on the notion of the accuracy of survey estimates. It is imperative to thoroughly understand and adjust for differentiation in data quality of underlying variables and constructs in order to validate internationally comparative research. It is therefore the aim of this paper to present a comprehensive analysis of this phenomenon and to critically evaluate the implications for comparative research.

Item nonresponse as survey error

The notion of item nonresponse is but a small piece in the broader discussion of survey errors. Survey errors derive from different sources: coverage, nonresponse, sampling and measurement (Groves, 1989: 11). While coverage, nonresponse and sampling errors occur because some units are not observed (nonobservation errors), measurement errors occur during the data collection period i.e. while units are being observed (observation errors). Item nonresponse errors are formally classified as the errors of nonobservation being one of two possible types of nonresponse error (the other one being the unit nonresponse). Respondent's inability or unwillingness to answer survey questions or simple omissions of them is treated as item nonresponse. Although technically the value of separate item is indeed not observed, the mechanics of the item nonresponse more closely resemble the measurement errors (Groves, 1989: 156). Furthermore, it manifests on and is influenced by much of the same factors as the other measurement errors. Main confounders of the item nonresponse are therefore the respondents' characteristics, the interviewer characteristics and the questionnaire as the mediator of the survey communications.

Determinants of item nonresponse

The idea of item nonresponse behaviour was first introduced by Ferber (1966), who studied the combined influence of respondent and question characteristics on item nonresponse. A few decades later, Andrews (1984) discussed the differential influence of various levels of the survey process on measurement errors through the notion of correlated measurement errors. More recently, the research expanded towards modelling interviewer effects and respondent characteristics within multilevel models (Pickery and Loosveldt, 1998; Hox et al., 1991). So far, three fundamental levels of item non-response correlates have been identified: respondent characteristics, interviewer characteristics and characteristics of survey questions.

Throughout the literature, item non-response most commonly relates merely to the characteristics of respondents. The main reason for this lies in the fact that these variables are usually very strong correlates (Francis & Busch, 1975; Grønhaug et al., 1988). Nevertheless, there are also situations where they are completely irrelevant (Goor & Verhage, 1999). Another reason for the predominant role of respondent's characteristics is that relevant data is more easily accessible, while an additional survey would be needed in order to obtain, for example, information on the interviewer's characteristics. Questionnaire characteristics, which are available in every survey, are also important correlates of item nonresponse. The relevance of the survey topic for the respondent is clearly a key determinant of item nonresponse (Kupek, 1998; Petty & Jarvis, 1996), and the most discussed aspects are the salience of the survey topic and its social desirability (Beatty & Herrmann, 2001). Question context is also an important determinant (Sudman et al., 1996: 116), particularly the impact of preceding questions, which can lead to item nonresponse. The initial purpose of such questions is to lower the respondent burden; however, the respondent can learn their meaning and use them as a tool to avoid unwelcome parts of the questionnaire (Knäuper, 1998). Both respondent and questionnaire levels of covariates of item nonresponse will be addressed in the proposed paper.

Data

The Generations and Gender Survey (GGS) is a panel survey of a nationally representative sample of 18-79 year-old resident population with at least three panel waves and an interval of three years between each wave (United Nations, 2005). As of to date 16 countries have collected Wave 1 data: Australia, Austria, Bulgaria, Czech Republic, Estonia, France, Georgia, Germany, Hungary, Italy, Japan, Lithuania, Netherlands, Norway, Romania and the Russian Federation. The first country to complete the data collection was Hungary in 2001. Belgium will become the 17th country with the GGS Wave 1 data and is currently in the field. Data were mainly collected via personal interview with (CAPI – Computer Assisted Personal Interviewing) or without (PAPI – Paper And Pencil Interviewing) computer assistance. Norway is the only country so far to use a mixed mode of data collection combining register information with CATI (Computer Assisted Telephone Interviewing) and drop-off-pick-up self-completion interviews. The completed Wave 1 sample sizes range from 7,682 respondents in Australia to 25,000 respondents in Hungary with the majority around the 10,000 respondent mark. Central survey topics are fertility, partnership, transition to adulthood, economic activity as well as the intergenerational and gender relations between people expressed in care relations or the organization of paid and unpaid work (Vikat et al. 2007) resulting in 1,700 variables per dataset on average.

Methods

A comprehensive analytical model is needed to simultaneously model item non-response at both respondent and questionnaire levels of the survey process. The major challenge for the analytical solution to the proposed models is the complexity of the data structure that does not correspond to the standard multilevel notation. The cross-classification of questions across respondents is not complete, since not all of the respondents are asked all of the questions due to the routing of the questionnaire. Filter questions determine the appropriate route for each respondent. The respondents answer a different number of questions, but the interviewer also poses a varying number of questions to a different number of respondents. Another characteristic of this specific design is the fact that at level one there is exactly one unit in each cell, since all questions are presented by one interviewer to one respondent only once. It is theoretically impossible to have more than one unit at the lowest level. In addition to these two data-structure problems, the estimation methods for standard cross-classified multilevel models are computationally very intense and cannot cope with large databases and a large number of explanatory variables (Hox, 2002: 137).

In order to address the above problems, the meta-analytical approach seems to provide the right answers. The approach is also known under the term Variance-known or V-known models (Raudenbush & Bryk, 2002: 206). Most commonly, the V-known models are used in the meta-analysis of published results (studies), where we have the estimate of a target parameter and its variance for each study. The analysis builds on the characteristics of the study (i.e. published research) as the explanatory variables of the variance in the estimates across all analysed studies. As suggested by Raudenbush and Bryk (2002: 207), the notion of the V-known models can be expanded beyond meta-analysis to any statistical coefficient in different settings. In our case, the survey questions stand as a different study to the analogy of a standard meta-analysis, while the effects of the respondents and interviewer together act as the corresponding target variable within each question (study). The unresolved structure of our data thus calls for a proxy solution to estimate the analytical model and each survey question is modelled separately in the first stage of our analysis. The need to use meta-analytical methods and the V-known models in the present analysis thus arises due to both the relatively large sample sizes and the initial false cross-classification matrix with an unclear data structure.

The two level models will thus be estimated in two stages. In the first stage, logistic regression models will be estimated, with response-nonresponse as the binary dependent variable and respondents' characteristics as explanatory variables. The results of these models (i.e. intercepts and their corresponding variances) will be pooled for the second stage, where question characteristics are used as level 2 predictors.

The item nonresponse models will be initially run on the pooled dataset of all eight countries. In the second stage of the analysis the models will be separately estimated for each country and findings will be compared. The results from both estimation stages will contribute to a better understanding of the survey response process and of the validity of international comparative research.

Bibiliography

Andrews, F.M. (1984). Construct Validity and Error Components of Survey Measures: A Structural Modelling Approach. *Public Opinion Quarterly*, Vol. 48, No. 2, pp. 409-442.

Beatty, P., **Herrmann**, D., (2002). To Answer or Not to Answer: Decision Processes Related to Survey Item Nonresponse. *In: R.M. Groves, D.A. Dillman, J.L. Eltinge, R.J.A. Little (eds). Survey Nonresponse.* John Wiley & Sons.

Francis, J.D., **Busch**, L. (1975). What we don't know about "I don't knows". *Public Opinion Quarterly*. Vol. 39, pp. 207-218.

Goor, H., van, **Verhage**, A.L. (1999) Nonresponse and Recall Errors in a Study of Absence because of Illness: An Analysis of Their Effects on Distributions and Relationships. *Quality & Quantity,* Vol. 33, pp. 411-428.

Grønhaug, K., **Gilly,** M.C., **Enis,** B.M. (1988). Exploring income non-response: A logit model analysis. *Journal of the Market research society.* Vol. 30, No. 3.

Groves, R.M. (1989). Survey Errors and Survey Costs. John Wiley & Sons.

Hox, J.J. (2002). Multilevel Analysis. Techniques and Applications. Mahwah, NJ: Lawrence Erlbaum Associates.

Hox, J.J., **Leeuw**, E.D., de, **Kreft**, I.G.G. (1991). The Effect of Interviewer and Respondent Characteristics on the Quality of Survey Data: A Multilevel Model. *In P. Biemer, R.M. Groves, L.E. Lyberg, N.A. Mathiowetz, S. Sudman (eds.). Measurement Errors in Surveys.* John Wiley & Sons. New York.

Knäuper, B. (1998). Filter Questions and Question Interpretation – Presuppositions at Work. *Public Opinion Quarterly*, Vol. 62, pp. 70-78.

Kupek, E. (1998). Determinants of item nonresponse in a large national sex survey. *Archives of Sexual Behavior*. Vol. 27, No. 6. New York.

Petty, R.E., **Jarvis**, W.B.G. (1996). An Individual Differences Perspective on Assessing Cognitive Processes. *In N. Schwarz, S. Sudman (eds.). Answering Questions: Methodology for Determining Cognitive and Communicative Processes in Survey Research*. Jossey-Bass Publishers, San Francisco.

Pickery, J., **Loosveldt**, G., (1998). The Impact of Respondent and Interviewer Characteristics on the Number of "No Opinion" Answers – A multilevel model for count data. *Quality & Quantity,* Vol. 32, pp. 31-45.

Raudenbush, S.W., **Bryk**, A.S. (2002). Hierarchical Linear Models. Applications and Data Analysis Methods. 2nd Edition. Sage Publications.

Saris, W. E., **Gallhofer**, I. N. (2007). Design, evaluation, and analysis of questionnaires for survey research. Wiley-Interscience.

Sudman, S., **Bradburn**, N.M., **Schwarz**, N. (1996). Thinking About Answers: The Application of Cognitive Processes to Survey Methodology. Jossey-Bass Publishers, San Francisco.

United Nations (2005) *Generations and Gender Programme – Survey Instruments*, United Nations, New York and Geneva.

Vikat A., **Spéder** Z., **Beets** G., **Billari** F. C., **Bühler** C., **Désesquelles** A., **Fokkema** T., **Hoem** J. M., **MacDonald** A., **Neyer** G. R., **Pailhé** A., **Pinnelli** A., **Solaz** A. (2007) Generations and Gender Survey: Concept and Design in *Generations & Gender Programme – Concepts and Guidelines*, United Nations, New York and Geneva.