



ESRA 2007 conference

25 – 29 June 2007 Prague, Czech Republic



European Survey research Association



ESRA conference 2007

The European Survey Research Association

June 8, 2007

Preface: ESRA activities 2005-2007

In 2005, the current committee started up the activities of ESRA. Our main activity was the conference in Barcelona, organised by ESADE and the Universitat Pompeu Fabra. In total 302 people from 26 different countries participated, and given the number of people participating and the variety of topics and the interesting discussions, we think that the conference was a big success. During the conference we held the first general assembly of the ESRA to determine whether we should establish an independent association or become a subsection of another association. It was decided at least for the time being to start our own organisation. A decision was also made with respect to the members of the board of the association and some other activities (see minutes of the meeting).

After the conference we applied for a grant of the ESF in order to evaluate the advantages and disadvantages of access panels which had led to a lot of discussions at the conference. Unfortunately this application has been rejected. This does not mean that no work was done in this context: the results of this work can be seen in the program of the conference in 2007.

In 2006 preparations started for an ESRA website and ESRA journal *Survey Research Methods* (SRM). The website has been developed and maintained by Daniel Oberski. The website contains relevant information about meetings and ESRA activities. This information is available for anybody but the members of ESRA will be notified if there is relevant new information.

SRM has been edited by Peter Lynn and Rainer Schnell. In January 2007 the first issue appeared and this month the second issue has appeared. The editors plan to provide 3 issues per year of the Journal. The journal freely available online for everybody. The members of ESRA will be notified when a new issue becomes available on the web.

The European Association for Methodology (EAM) organized its biannual conference in Budapest. At that meeting ESRA was well represented with several different sessions of its members.

In 2007 we formally established the association by registering the ESRA in Barcelona. The constitution has been available for some time at the ESRA website and should be discussed and formally accepted at the general assembly during the ESRA conference in Prague 2007.

In 2007 the bulk of the work was the organisation of the second ESRA conference. We were very pleased with the offer of the University of Economics and the Czech Statistical office to organize the conference in Prague. Especially Martin Zeleny has done a lot of work for the local organisation of the conference. Daniel Oberski, Stefanie Huber, and I have been actively trying to collect all the papers that were sent to us and to organise the program in an understandable way. We hope that we were successful in this matter. Our work has not been easy. We received even more papers than last time and the variety of topics was also larger. We are very grateful for your enthusiasm for the conference. We hope that you have also expressed this interest in our activities by becoming member of ESRA .It does not cost much but it will support us in our activities during the next two years.

On behalf of the committee of ESRA I wish you an interesting conference.

Willem E. Saris,
Chairman of ESRA

<http://esra.sqp.nl>
<http://surveymethods.org>

Special Events

Besides the different sessions, several special events are also planned during the conference this year. Below you will find a short description of each event; please check the time schedule for more information on the location and the definite time. Besides being distributed in paper form, the time schedule is also available online at <http://esra.sqp.nl/esra/conferences/2007>.

Welcome reception

Opening of the conference

Monday June 25 17:00

Welcome to the conference by Jakub Fischer, Vice-Dean for Science and Research, Faculty of Informatics and Statistics and Martin Zeleny, Dept. of Economic Statistics, Faculty of Informatics and Statistics of the University of Economics, Prague, followed by an opening speech by Willem Saris of ESADE Barcelona and chairman of ESRA.

Keynote speeches

Measuring Social Capital

*Jan van Deth; University of Mannheim, Germany
Monday June 25 from 17:15–18:00*

The Problems of Testing Structural Equation Models

*Albert Satorra; Universitat Pompeu Fabra, Spain
Wednesday June 27 from 18:00–18:45*

Structural Equation Models (SEM) are often used in many fields of research. Although the testing of such models is sometimes taken for granted, those who dare question it risk opening up a Pandora's box filled with furious discussions and complex statistics.

However, the asymptotic theory underlying SEM model tests need only be considered from two perspectives: model misspecification and the violation of assumptions.

This presentation discusses the power of the test to detect model misspecifications, the influence on testing decisions of violations of the assumption of normality, and the interaction between the two. It is shown that on the one hand, testing is robust to non-normality if a well-known scaled test statistic is used. On the other hand, the non-normality of the data affects the power to detect misspecifications of the model. The power of the test to detect given misspecifications can be computed regardless of non-normality.

Web Surveys: advantages and disadvantages

*Vasja Vehovar; University of Ljubljana, Slovenia
Friday June 29 from 15:45–16:30*

The use of the World Wide Web for survey data collection has been increasing already for a decade. There are several research fields where Web surveys are used regularly because coverage presents a smaller problem and probability sampling is possible – examples are establishment surveys, consumer research of online services, research on students and professionals, etcetera.

But web surveys are also regularly used in the field of health research or for market research where researchers must content themselves with non-probability samples. While web surveys are already widely used, survey methodologists still try to show whether they are actually as valid and reliable as more conventional survey methods.

One approach to do this is by comparing the results of web surveys to the results of more conventional survey methods such as a mail survey, telephone survey, or face-to-face survey on a particular topic. This invited paper will thus give an overview of the type of research designs used for this purpose, issues addressed (e.g. data quality, measured with differences in unit and item-nonresponse, consistency of answers, substantive answers etc.; cost issues), and results revealed.

Book exhibition

Tuesday June 26 – Friday June 29

During the entire conference, an exhibition with books of possible interest to survey researchers is open for all participants.

Book presentation: Measuring Attitudes Cross-Nationally

Tuesday June 26 at 19:30

Sage Publications invites all ESRA delegates for a glass of wine in the Foyer at the end of Tuesday's sessions at 19.30. The purpose is to celebrate the recent publication of 'Measuring Attitudes Cross-Nationally: Lessons from the European Social Survey', edited by Roger Jowell, Caroline Roberts, Rory Fitzgerald and Gillian Eva. This is the first book to describe the methodology and approach of the Descartes Prize winning European Social Survey, with which many of you will be familiar.

General assembly of the ESRA

Friday June 29 at 16:30

All members of the ESRA are invited to partake in the second general assembly of the organisation, to be held just after the keynote speech by Vasja Vehovar and just before the closing of the conference in the main hall.

Agenda

1. Opening of the meeting
2. Minutes of the meeting in Barcelona
3. Overview of the activities of ESRA 2005-2007
 - a) Conference in Barcelona
 - b) Constitution of ESRA

- c) Official registration of the association
 - d) ESRA website
 - e) Journal SRM
 - f) Rejected ESF application concerning access panels
4. Financial report over 2005-2007
 5. Activities of ESRA 2007 - 2009
 - a) ESRA web site
 - b) Journal SRM
 - c) Participation in the conference in Berlin
 - d) Organization of the next conference
 6. Budget for 2007-2009
 7. Membership fees
 8. Place for the conference 2009
 9. Membership of the committee of ESRA
 10. Other possible issues

Closing speech

Friday June 29 at 17:45

Conference dinner on the Moldau

Friday June 29 at 19:45

The conference will be ended with a dinner for participants, which will take place on a boat taking us on the Moldau. Those who registered for the dinner are kindly asked to be on time.

Topics and sessions

Topics

Comparative Research	7
Quality of Survey Questions	41
Measures for Political Science Research	75
Data Analysis	99
Sampling and Nonresponse	131
Mode of Data Collection and Data Enhancement	157
Specific Topics I	189
Specific Topics II	225

Overview

1	Comparative Research	7
1.1	Comparative Pretesting	8
1.2	Comparative Questionnaire Design	12
1.3	Harmonisation of Survey Data	17
1.4	Data Collection in Cross-cultural and Cross-national Surveys	21
1.5	Sampling for Cross-national and Multi-frame Surveys	25
1.6	Methodological Issues in Multilevel Analysis for Cross-national Research	28
1.7	Measurement Quality of Concepts in the ESS	32
1.8	Developing Attitudinal Indicators	36
2	Quality of Survey Questions	41
2.1	Software Presentations	42
2.2	Questionnaire Translation and Adaptation	44
2.3	Questionnaire Development and Testing	50
2.4	Communicating Survey Quality to Users	56
2.5	Interviewer and Respondent Behaviour in Survey Interviews	59
2.6	Augmenting Survey Data by Bio-indicators, Geo-data and Record-linkage	63
2.7	Improving the Validity of Psychological Questionnaires in the Context of Survey-based Research	65
2.8	Enhancing Survey Methodology with Qualitative Methods	70
3	Measures for Political Science Research	75
3.1	Analysing Attitudes Towards Migration with Large Comparative Survey Data	76
3.2	Measuring the Quality of Democracy on the Basis of Survey Data	79
3.3	Human Values	83
3.4	Measuring and Explaining Trust	87
3.5	European Values Study	93
4	Data Analysis	99
4.1	Analysis Strategies for Cross-cultural Research	100
4.2	Analysing and Estimating Reciprocal Relationships in Panel Data by Means of Continuous Time Methods	104
4.3	Applied Missing Data Strategies in Survey Validation and Use	108
4.4	Small Area Estimation: Theory and Applications	112
4.5	Categorical Data Arising from Dependent Observations: Multilevel And/or Marginal Models	116
4.6	Models of Analysis for Defining and Constructing Subjective Data and Indicators	119
4.7	Nonresponse and Measurement Error	124
4.8	Testing Structural Equation Models	127

5	Sampling and Nonresponse	131
5.1	Miscellaneous	132
5.2	Reducing, Detecting, and Adjusting for Nonresponse Bias	139
5.3	How to Cope with Decreasing Response Rates	144
5.4	Survey Participation and Non-response Bias in Surveys and Access Panels	146
5.5	Selection Bias in Panel Research	149
5.6	Probability-based Recruitment for Online Panels	152
6	Mode of Data Collection and Data Enhancement	157
6.1	Perils and Promises of Web Surveys	158
6.2	From Questionnaires to Data File: the Role of the Data Producers and Disseminators	162
6.3	The Use of Paradata for Evaluating and Improving Survey Quality	169
6.4	Using Contact Forms for Data Quality Assessment of Cross-national Surveys	175
6.5	Mobile Phone Usage in Survey Research: Implications for Data Quality	178
6.6	Mixing Modes of Data Collection	183
7	Specific Topics I	189
7.1	Emotional Intelligence Research	190
7.2	Demographic and Socioeconomic Variables in Cross-national Surveys	193
7.3	Cost-effective Data Editing for Business Surveys	198
7.4	The Impact of Media-reported Events on Attitudes and Opinions: Measurement and Analysis	201
7.5	Network Survey Measurement	207
7.6	Statistics on Income and Living Conditions (EU-SILC)	212
7.7	Income Measurement in Surveys	215
7.8	Survey Measurement of Service Quality	221
8	Specific Topics II	225
8.1	Ethical Considerations in Survey Research	226
8.2	Secularisation Under Different Regimes	230
8.3	Health Surveys in Europe	233
8.4	The Stability of Political Opinions & Attitudes	236
8.5	Time Use Surveys	240
8.6	Measuring and Explaining the Fear of Crime	247
8.7	Survey Data and Public Dissemination	250
8.8	Survey Evaluation Methods Using Behaviour Coding and Sequence Analysis	254
8.9	Difficult Groups in Survey Research	257
	Index of people	263

1 Comparative Research

Sessions in Comparative Research

Comparative Pretesting	8
Comparative Questionnaire Design	12
Harmonisation of Survey Data	17
Data Collection in Cross-cultural and Cross-national Surveys	21
Sampling for Cross-national and Multi-frame Surveys	25
Methodological Issues in Multilevel Analysis for Cross-national Research	28
Measurement Quality of Concepts in the ESS	32
Developing Attitudinal Indicators	36

1.1 Comparative Pretesting

Kirsten Miller; Centers for Disease Control and Prevention, United States

Gordon Willis; National Institutes of Health (NIH), United States

Presentations

1. *Introduction to the Session* – G. Willis (United States)
2. *Design and Analysis of Cognitive Interviews for Cross-national Testing* – K. Miller (United States)
3. *The Use of Cognitive Interviews to Assess Construct Overlap in Cross Cultural Surveys* – J. Padilla Garcia (Spain)
4. *Measuring Employers Social Contributions in the Context of EU-SILC: Results of a Two-stage Pretest for Germany* – B. Gauckler (Germany)
5. *Implementation and Results of a Cross-national, Structured-interview Cognitive Test* – J. Madans (United States), K. Miller (United States), A. Maitland (United States)

Presentations in the session “Comparative Pretesting”

Introduction to the Session

Gordon Willis; National Institutes of Health (NIH), United States

These researchers will focus on a problem that has existed for some time, but that is increasingly important as survey organizations conduct multi-ethnic, multi-cultural, or cross-national surveys: How can survey questions, questionnaires, and survey administration in general be designed so that the obtained results are conceptually and analytically equivalent? In particular, what processes and procedures can we use to establish that the principle of cross-cultural comparability holds? The current papers approach the problem from an empirical point of view, presuming that testing prior to field administration, in large part inspired by the practice of cognitive interviewing, can be useful in this regard. As such, we will hear about the methodological directions taken within several of these investigations, along with substantive results. Following presentations by the authors, I will further discuss a few points made by each, with some summary remarks concerning potential next steps in this vital research area.

Design and Analysis of Cognitive Interviews for Cross-national Testing

Kirsten Miller; Centers for Disease Control and Prevention, United States

The purpose of this paper is to discuss the promise of traditional cognitive testing in identifying comparability problems in questions intended for multi-national surveys. The paper is based on a multi-national cognitive testing project to test health status questions developed from the Budapest Initiative, a UNECE/WHO/Eurostat task force on the measurements of health states. Because cognitive testing directly examines the thought processes that respondents use to answer survey questions, including their interpretations of those questions, this method holds much promise for uncovering culture or language-related problems. Nevertheless, because of specific characteristics inherent to the methodology (such as small sample size, non-standardized interviewing protocols and differing criteria for what constitutes a test finding), the method has potentially serious limitations that must first be addressed. Drawing upon the work performed by the US for the Budapest Initiative, this paper will 1) outline those characteristics, 2) pose design and analysis strategies for overcoming those limitations and 3) illustrate the types of culture-specific findings that can be identified using the method.

The Use of Cognitive Interviews to Assess Construct Overlap in Cross Cultural Surveys

Jose Luis Padilla Garcia; University of Granada, Spain

The validity of comparisons between the measures obtained in cross-cultural surveys depends on the level of equivalence achieved. The equivalence among measures relies on the extent of construct overlap in the language and cultural groups. Cognitive pretesting methods can help to improve different aspects of translated versions of questionnaires in cross-national and cross-cultural surveys. The aim of the present study is to explore the usefulness of cognitive interviews for gathering evidence of the extent of construct overlap measured by questionnaires between different linguistic and cultural groups. The most significant results of cognitive pretesting of the Spanish version of the European Health Status Module (EHSM) are presented. The EHSM is one of the four modules that will constitute the future European Health Interview Survey. The Spanish version of the EHSM is the product of a design of direct translation starting from the source version in English. The target questions analysed during cognitive pretest were identified taking account of linguistic and cultural issues. One of the distinctive characteristics of the cognitive pre-test carried out is the attention paid to disabled people. The total number of participants in the cognitive interviews was 25 people. People with sensory and mobility disabilities took part in the cognitive interviews also. 12 target questions were analysed by means of cognitive interviewing. The general pattern of results allows us to be confident that the Spanish version

of the questions is functionally equivalent to the English version. The effectiveness of the cognitive interview depends on having a list of target behaviours. The inclusion of disabled people has allowed us to analyse the interpretation of questions and key terms such as “health in general”, “long-standing illness or health problem”, “being limited”, etc., in contrast with the interpretations of non-disabled people. Lastly, methodological proposals for improving the use of cognitive interviewing, for instance, the effects of different types of follow-up probes and the adaptation of follow-up probes for disabled peoples, are discussed.

Measuring Employers Social Contributions in the Context of EU-SILC: Results of a Two-stage Pretest for Germany

Britta Gauckler; Federal Statistical Office, Germany

The recommendations of the ‘Canberra Group’ (Expert Group on Household Income Statistics) state that social employer’s contributions are a relevant part of employees’ income. Consequently, the European Survey on Income and Living Conditions (EU-SILC) requests the measurement of this income component in a separate target variable. The measurement will be mandatory for all participating countries, if studies show the feasibility. This precondition indicates the difficulties to implement EU recommendations for each country due to specific cultural and political systems.

For the German EU-SILC operation there are two main complications connected to a possible measurement of employers’ social contribution. Firstly, the system of employer’s social contributions in Germany is divided into two parts: social contributions which are mandatory by law and those which are voluntary (or compulsory by contract). This mix makes the measurement a complex task. Secondly, due to data protection the information of employer’s contribution has to be collected via the employee. Therefore, the two leading questions for the study were a), whether employer’s social contributions can be measured within a postal survey via employee information and b), due to the complex system of employer’s social contributions in Germany, if respondents are able and willing to provide as demanded.

In order to respond to the requirements of the study a two-stage pretest approach was chosen for the evaluation of a possible implementation. The first, qualitative part was particularly used to identify possible problems of the question-and-answer-process. For the study 15 cognitive interviews via a standardised compendium were conducted. The main used technique within the interviews was probing. The second, quantitative field test could clarify whether the revised instrument based on the qualitative part enables the collection of valid and reliable data for Germany. On basis of the first qualitative results the questionnaire underwent revisions and was extended with debriefing questions at the end of the questionnaire. The quantitative field test with 599 respondents (net) was conducted in August and September 2006.

The instruments used for the feasibility study - the combination of qualitative and quantitative methods- in particular cognitive interviews in combination with the field test on a quantitative basis, was well suited for the analysis of a topic such as the employer’s social contributions. The recognized shortcomings of the qualitative testing had been (unfortunately) confirmed by quantitative testing, even though a revised and improved questionnaire was used. The results show that the data collection of employer’s social insurance contributions via an employee survey is problematic in the case of Germany. In summary, the target variable on employers’ social contribution can only partly be integrated in EU-SILC for Germany.

Implementation and Results of a Cross-national, Structured-interview Cognitive Test

Jennifer Madans; Centers for Disease Control and Prevention, United States

Kirsten Miller; Centers for Disease Control and Prevention, United States

Aaron Maitland; Centers for Disease Control and Prevention, United States

This paper describes the Washington Group project to test a short battery of disability questions developed for national censuses. The study used an unusually structured cognitive test protocol and

was administered to a total of 1,287 respondents selected from convenience samples in fifteen countries in Central and South America, Asia and Africa. The test protocol consists of the six core disability questions followed by probe questions that were designed to illustrate: 1) whether core questions were administered with relative ease; 2) how core questions were interpreted by respondents; 3) the factors considered by respondents when forming answers to core questions; and 4) the degree of consistency between core question responses and responses to a set of more detailed questions addressing respondent functioning in each of the domains covered by the six core questions. Demographic and general health sections are also included to examine whether test questions perform consistently across all respondents, or if nationality, education, gender or socio-economic status impact the ways in which respondents interpret or consider each question.

The paper will first articulate the analytic objectives of the cognitive test, the type of data required to perform the test as well as the challenges to collecting such data among disparate economic and cultural contexts. Primary discussion, however, will focus on the test results, illustrating the usefulness of this method in question evaluation.

1.2 Comparative Questionnaire Design

Session coordinated by:

Brad Edwards; Westat, United States

Sue Ellen Hansen; University of Michigan, United States

Presentations

1. *Differences in Cognitive and Communicational Styles Across Cultures and Effect on Survey Measurement Error* – E. Peytcheva (United States)
2. *Context Effects in Surveys Used in Cross-cultural Research: Implications of Individualism and Collectivism* – A. Uskül (United Kingdom)
3. *Improving the Documentation of (cross-national) Questionnaire Development* – R. Fitzgerald (United Kingdom)
4. *Dealing with Divergent Demands: Questionnaire Construction Between Longitudinal and Comparative Constraints* – M. de Groot (Netherlands)
5. *Comparative Questionnaire Design: a Review of Current Practice* – S. Hansen (United States), J. Harkness (Germany), A. Villar (United States), M. de la Puente (United States)
6. *Survey Instrument Design for Comparing Groups: an Overview* – B. Edwards (United States)

Presentations in the session “Comparative Questionnaire Design”

Differences in Cognitive and Communicational Styles Across Cultures and Effect on Survey Measurement Error

Emilia Peytcheva; University of Michigan, United States

Surveys in multiple countries are becoming increasingly common, often requiring administration in multiple languages. The desire to compare respondents in different countries has led to studies focusing on how to make the questionnaires more similar, both through translation and item/scale equivalence. However, little attention has been paid to differences in cognition and communication across cultures in the context of social surveys and the consequences of how respondents in different cultures may process the same set of survey stimuli.

Recent research in cognitive psychology has demonstrated persistent differences between Westerners and East Asians on a variety of cognitive tasks. East Asians have been found to exhibit greater sensitivity to context, background details and relationships among background elements. Focusing on a large portion of one’s surroundings and paying attention to objects and background and relationships between objects and background influences the understanding about the complexity of the world. Asians and Westerners have been shown to have different dominating perceptions of change and stability. Specifically, Asians perceive the world as constantly changing in a cyclical fashion, while Westerners believe in the stability of what is surrounding them and if a change occurs, it follows a linear, predictable pattern.

Differences between Westerners and East Asians have also been found in self perception and relationships with others. Specifically, Western cultures promote the independent self-construal and perceive the self distinct from others, possessing unique abilities, thoughts and feelings. In contrast, East Asian cultures promote the interdependent self-construal and conceptualize the self as defined and connected by others, others’ feelings, thoughts and social roles. Interdependent societies (or “high context societies”) highly value the maintenance of social harmony and avoidance of conflict, thus put an emphasis on ‘reading between the lines’ or ‘reading the mind’ of the others.

Such cognitive and communicational differences can induce culturally specific context effects. For example, considering the cyclical vs. linear way of thinking about change, a report by an East Asian about a current state can be affected by the response provided in the preceding questions about a past state, but no such effect will be observed for Western respondents. Similarly, given that interdependent cultures are sensitive to the conversational common ground, it will be expected that they will attend closely to the content and presentation of survey questions and disregard information they had provided in earlier questions.

Data from the Gallup World Poll are used to test the existence of such cultural differences in processing. The main focus of discussion is the contrast in cognitive and communication styles between respondents from Japan and the US in terms of responses to series of general-specific questions and questions about past, current and future states. Additional analyses test to what extent the results can be generalized further to countries with predominantly independent vs. interdependent worldviews. The findings have implications for the development of functionally equivalent questionnaires across cultures.

Context Effects in Surveys Used in Cross-cultural Research: Implications of Individualism and Collectivism

Ayşe Uskül; University of Essex, United Kingdom

Psychological research provides evidence for systematic cross-cultural differences in psychological processes such as level of sensitivity to social context, meanings inferred from communication, and what is likely to be salient in memory, which are expected to lead to cross-cultural variation in processes underlying survey responding. We have recently theorized that systematic differences in survey responding

should be observed between members of cultural groups that differentially endorse an individualistic versus a collectivistic cultural frame that guides the experience of the self and the social world (Uskul, & Oyserman, 2006). In an attempt to test this theoretical framework, we first hypothesized greater sensitivity to survey question context among members of collectivistic cultures than members of individualistic cultures. Second, we hypothesized that differences in context effects should depend on question content such that responses given to questions that are culturally-relevant should be influenced by question context to a greater extent than questions that are culturally-irrelevant.

We tested these hypotheses by examining the effects of two scales, one ranging from 0 to 10 (unipolar), the other from -5 to +5 (bipolar), in an individualistic (European-American) and a collectivistic group (Turkish). In Study 1, questions concerned evaluation of oneself on a list of attributes such as successful, fair etc. Findings replicate the pattern observed by Schwarz and colleagues (Schwarz et al., 1991); responses have a lower mean on the unipolar than on the bipolar scale suggesting a lack of preference for the lower end of the bipolar scale which conveys the opposite meaning of the attribute on which participants rated themselves (e.g. success \neq failure). Findings also reveal, confirming the first hypothesis, an effect of culture; change in scale context led to a significantly larger difference in the means among Turks than among Americans due to Turks' lower likelihood of using the upper end of the unipolar scale than Americans'.

In Study 2, we tested whether the observed findings in Study 1 would generalize to different question content. This time, participants evaluated their parents on the same list of attributes as well as people at the age of their parents. American participants showed the expected scale effect regardless of the type of group evaluated. For Turkish participants, however, a clear ingroup-outgroup distinction emerged, such that only parent ratings were subject to the scale effect; in evaluation of people at their parents' age – a culturally irrelevant group –, participants were as likely to use the lower end of the bipolar scale as they did of the unipolar scale.

These studies provide a systematic test of response processes underlying self-report in a survey context across cultural groups that differentially endorse individualism and collectivism and point to important interactions between cultural background, question context, and question content. Findings suggest that plain interpretations of mean differences derived from self-report measures commonly employed in cross-cultural research may lead to unwarranted conclusions.

Improving the Documentation of (cross-national) Questionnaire Development

Rory Fitzgerald; City University, United Kingdom

Despite increases in the transparency of elements of survey design in recent years, the process of questionnaire development has somehow bucked this trend. True, we are often told in technical reports that qualitative work played a role in the design of new questions, and that there was a pilot of a given size, and that certain scales were developed after extensive iteration and analysis. But the detail about why certain new questions took the form they did, or why certain well-tried questions from other surveys were preferred to others usually remains a mystery. In a cross-national survey this is compounded because translators often have difficulty understanding the purpose of the original source questions.

In the European Social Survey (ESS), we have tried to rectify such omissions. The ESS questionnaire consists of two main sections. A 'core' section comprising about half of the questionnaire contains largely stable questions that underwent testing and evaluation well before Round 1 of the survey went into the field and have since been repeated in each biennial round. The rationale for and development of these questions was described and explained in an extensive report. The other half of the questionnaire consists of 'rotating modules' on specific topics awarded to multinational teams. The modules go through a protracted period of design and development designed to achieve reliability and validity, as well as attaining equivalence across countries and languages.

The process combines expert review, extensive consultation across countries, and an unusually detailed set of translation protocols. It also incorporates statistical modelling that assesses the reliability and validity of survey items and scales (prior to and after piloting) and investigates the latent structure

of underlying concepts. Based on this input and the advice of substantive and methodological experts, items or scales are then adopted, re-designed or discarded.

The work on the rotating modules is carried out jointly by the selected team of academics in collaboration with a subset of the CCT. To some extent during earlier rounds of the project and as far as possible during the latest Round 3, attempts have been made to better document the process. But further work is needed to improve the structure of and accessibility to this documentation.

Data users are already provided with thorough documentation of the ESS methodology and questionnaire design needs to meet these same high standards. During Round 4 we intend to improve such documentation by asking the Question Design Teams and CCT to document process via a pre-designed template. The template and guidelines will be included in the paper and suggestions for improvements discussed. The paper will consider the balance between the burden of demands on the design team and the utility of the documentation for translators and data users. The ability to use this style of documentation for other surveys will also be discussed in the context of improving comparability of questionnaire design in a cross-national context.

Dealing with Divergent Demands: Questionnaire Construction Between Longitudinal and Comparative Constraints

Merel S. de Groot; University of Twente, Netherlands

In 2002 a new Local Government Act was introduced in The Netherlands. Initiators of this new Act expected attitudinal and behavioural change of municipal councillors. To determine if these expectations come out, a pre-measurement and post-measurement are needed. In 1999 a pre-measurement was conducted, a post-measurement will be conducted in the spring of 2007. With respect to this post-measurement several aspects need to be thought over, a.o. data collection, research population, research sample and the questionnaire. Especially the questionnaire construction faces various difficulties, resulting from divergent demands imposed by longitudinal versus internationally comparative comparability. To start it is a longitudinal research: the questionnaire should be comparable to the 1999 questionnaire. Secondly, the Dutch research is imbedded in an international council survey conducted in 15 European countries (Municipal Assemblies in European Local Government in Transition: MAELG project). Each participating country has one or more national coordinators that will conduct the survey. The coordinators/experts have been working together on the development of the survey design, the questionnaire and strict guidelines on issues such as sampling and response rates. The aim is to achieve the highest standards of cross-national research. Both objectives – longitudinal and internationally comparative comparability - ask for several requirements, sometimes coinciding, sometimes conflicting. This paper sets forth a number of questionnaire formulation problems stemming from divergent demands, and argues the solutions we have chosen during the preparation on the council survey. The problems and solutions are embedded in a more general framework of identical/equivalent indicators (Van Deth (ed.) 1998; Przeworski and Teune 1966/1970).

Comparative Questionnaire Design: a Review of Current Practice

Sue Ellen Hansen; University of Michigan, United States

Janet Harkness; Centre for Survey Research and Methodology (GESIS-ZUMA), Germany

Ana Villar; University of Nebraska, Lincoln, United States

Manuel de la Puente; US Census Bureau, United States

Design of questionnaires for comparative research generally involves the following four decisions:

1. whether to adopt existing questions, adapt existing questions, write new questions, or develop a questionnaire that is a mix of adopted, adapted, and new questions;
2. whether to ask the same questions (ASQ), ask different questions (ADQ), or develop a questionnaire that is a mix of the same and different questions;

3. in source questionnaire and translation models, how much emphasis to place on actively ensuring there is cross-cultural input in the various stages of developing a source questionnaire and in the testing of the source questionnaire and decisions about potential adaptation needs, rather than only turning to cross-cultural aspects when version translation begins; and
4. how and when testing of questions produced is undertaken.

The literature sometimes speaks of simultaneous, parallel and sequential designs in this respect – simultaneous approaches attempting to make constant cross-cultural input, parallel designs also taking into consideration the cross-cultural requirements at the questionnaire design and testing stages, but in less intensive fashion, and sequential models addressing cross-cultural requirements mainly at the stage of translation.

Each decision represents a continuum in terms cultural input involved in design, and the associated cost, resources, and time required for development. At one extreme is a lower cost sequential development approach that aims to keep things as similar as possible across cultures on the basis of language. Here questions from other studies are often used, and replicated verbatim, in close translation or modified if necessary. At the other extreme lie higher-cost simultaneous development approaches. These can be used to develop both ASQ or ADQ questionnaires in a variety of ways.

Questionnaires developed for cross-cultural research are not automatically pretested and when pretesting does occur, new challenges are faced.

This paper reviews current practice in the design and evaluation of questions for comparative research, drawing on research in a variety of disciplines, including quality of life research, health and mental health, marketing research, and survey methodology. It will classify and summarize design practice broadly, including approaches to identification of underlying constructs, dimensions and indicators, and formulation of questions and question wording.

Survey Instrument Design for Comparing Groups: an Overview

Brad Edwards; Westat, United States

Questionnaires for survey interviews occupy an interesting space between scientific instruments and human discourse. They provide a standard text for interviewer questions, standardized ways to record respondent answers to questions, and a navigation map through the interview. But close observation of interviewer-respondent interactions during an interview typically reveal a number of slips from the script. Often these slips represent important aspects of the interaction as conversation: requests for clarification or other help, repairs to errors of understanding, suggestions that might keep the conversation from going astray, etc. Questionnaires are created as neutral, scientific instruments by researchers. They are products of the researchers' best effort to define the constructs for measurement and to operationalize these constructs in a specific time and culture so they can serve as a uniform stimulus and be understood in the same way by all or almost all members of the study population. But the researcher is not an actor in the survey interview; the interviewer and respondent are left to sort out their own interpretation of the survey questions.

Comparing groups is a common survey goal. Examining differences in frequencies by gender, by age group, by education or employment status, by income, or by race or ethnicity are often a first step in analysis of survey data. But all respondents (and interviewers) infer the meaning of questions through their own cultural "filters." If these filters produce different understandings of a given question for different groups in the study population, group comparisons become questionable.

This paper provides an overview of tools available to the survey designer for determining whether a question is appropriate for examining group differences, and introduces issues that can be especially problematic in comparative surveys across cultures, languages, and nations. It also serves as an introduction to the papers that follow in this session, on: basic issues and current practice in comparative questionnaire design; cultural differences in cognitive and communication styles; context effects arising from individualism and collectivism; comparative issues in longitudinal designs; and documentation problems in cross-national questionnaire development.

1.3 Harmonisation of Survey Data

Session coordinated by:

Peter Granda; University of Michigan, United States

Christof Wolf; Centre for Survey Research and Methodology (GESIS-ZUMA), Germany

Presentations

1. *Harmonizing Ageing Surveys* – J. Lee (United States)
2. *Introduction to the Session Harmonisation of Survey Data* – P. Granda (United States)
3. *Conceptionalising Supervisory Status As a Core Variable in Cross-national Surveys* – H. Wirth (Germany), R. Pollak (Germany), W. Müller (Germany), F. Weiss (Germany), G. Bauer (Germany)
4. *Production of the Harmonised European Time Use Database* – I. Niemi (Finland), H. Pääkkönen (Finland), P. Hyytiäinen (Finland), P. Väisänen (Finland)
5. *Harmonization of the Collaborative Psychiatric Epidemiology Surveys (CPES)*. – S. Hansen (United States)

Presentations in the session “Harmonisation of Survey Data”

Harmonizing Ageing Surveys

Jinkook Lee; Ohio State University, United States

One of the goals of the recently collected international data such as English Longitudinal Study of Ageing (ELSA), Survey of Health, Ageing, and Retirement in Europe (SHARE), and Korean Longitudinal Study of Aging (KLoSA) that have been modelled after the Health and Retirement Study (HRS) is to allow for comparative international studies. Such comparative studies require harmonization of key measures in common amongst the data sets. Potential benefits of cross-national studies of ageing are beginning to become widely recognized, and researchers have produced insightful findings from such studies. However, such studies are still small in number, and scientific opportunities have not been fully exploited yet.

Partly this is due to the fact that it is not straightforward to know which measures in the various surveys are fully similar and which are not. Moreover several countries are in the process of developing similar data sets, including China and Japan. Without wanting to impose harmonization on these new efforts, international comparative research would clearly benefit if new efforts would learn from the existing surveys and would be encouraged to harmonize their instruments with what is already available in other countries. In many cases, individual countries have selected non-harmonized measures simply because it is too complex and difficult to know what the existing measures in the current international surveys are.

In this light, I provide harmonized information, presenting a systematic comparison of the survey instruments and document the equivalence and differences across the surveys. A systematic examination of these large-scale ageing surveys creates efficiencies for scientists who are interested in conducting cross-national studies of ageing. Given the large amount invested in collecting each of the ageing surveys, the benefits of this systematic comparison of existing surveys greatly exceed its costs.

In examining comparability, it should be noted that in certain domains data inherently exhibit substantial country-specific heterogeneity. For example, it is impossible to generate compatible raw data on welfare programs, as different countries have organized these differently. Different countries also have different public and private pensions, educational systems, and different financial products and institutions, so for such data, only ex post comparability can be established. On the other hand, for most demographic characteristics, health events, and expectations, ex ante comparability can be established if the survey instruments are equivalent. Therefore, acknowledging such differences in comparability, I evaluated the cross-national comparability of the ageing surveys. Specifically, in determining the practical comparability, the following tasks were performed:

1. To compare concepts a survey is intended to measure,
2. To compare survey questions,
3. To compare the range and wording of the allowed responses,
4. To identify the criteria for determining which respondents would answer which questions,
5. To compare the time period over which respondents report retrospective data,
6. To examine the skip logic of the questionnaires, and
7. To create aggregates of variables necessary for ex post harmonization

Introduction to the Session Harmonisation of Survey Data

Peter Granda; University of Michigan, United States

This brief introduction to the session on harmonisation (no more than 10 minutes) will attempt to present this topic in the context of a world of survey data that now extends beyond a half century of data collection and analysis. The simple question of what constitutes harmonised data in itself might engender lively and diverse replies: cross-national surveys all intended to be comparable in both design, collection, and data file production; similar surveys done in several countries which may become candidates for harmonisation after the fact; and even cross-sectional surveys done in a single country that focus on a particular topic that researchers and archives want to combine to produce a new data resource for the scholarly community.

How far does one go along this path? Some variables across surveys may be compared while others could be fully harmonized. The context in which each variable exists in the survey must also be considered. Weighting, variance estimation, and dealing with item missing values all may present significant additional barriers to the harmonisation process. Data producers must thoroughly understand the sampling frames and other design characteristics of all component surveys as well.

How in the end are all of these decisions presented to interested users? The documentation burden is equally great because it is imperative that these users know what they can and cannot do with the new data product. How also will the new product become available? What role will the social science data archives play in the dissemination of information about harmonisation as well as in the distribution of the data and documentation itself?

These are some of the questions which data producers and archives must face before harmonisation projects can move forward.

Conceptualising Supervisory Status As a Core Variable in Cross-national Surveys

Heike Wirth; Centre for Survey Research and Methodology (GESIS-ZUMA), Germany

Reinhard Pollak; Wissenschaftszentrum Berlin für Sozialforschung (WZB), Germany

Walter Müller; University of Mannheim, Germany

Felix Weiss; University of Mannheim, Germany

Gerrit Bauer; University of Mannheim, Germany

With the increasing availability of cross-national data and research there is also an increasing need for internationally comparable social classifications. One important precondition for achieving a harmonized measurement is the comparability of the core variables used for operationalisation. The focus of our paper is exactly on this problem. One important core variable in the operational definition of the new European-Socioeconomic-Classification (ESeC) is the so-called supervisory status. The position of supervisors are seen somewhere between managers and ordinary employees. Since there are different degrees of supervisor responsibility throughout the workforce, it is a notorious problem on which basis and where exactly the boundaries should be set. In our paper we examine this problem in two ways. First, we explore the procedures used to measure the supervisory status in the European Social Survey (ESS) and the Labour Force Survey (LFS). Both of which are major European Surveys aiming at providing data with a high degree of cross-national comparability. Second, we present findings based on a pilot study that examines to which proportion employees are identified as supervisors depending on the operational procedure. In this study, we asked two randomly selected subsamples of a telephone survey for their supervisor responsibility, including the questions asked in the ESS, the LFS in Great Britain and the LFS in Germany. The findings indicate that depending on the operationalisation, the groups identified as supervisors not only differ in their size and the extent of their supervisor responsibilities but also lead to quite substantial variation in distribution of the ESeC class-scheme. Our conclusion is, that it would be highly useful if a standard measure of the supervisory status could be developed and if this standard measure applied in uniform ways in future scientific surveys and the collection of official data.

Reinhard Pollak, Heike Wirth, Felix Weiss, Gerrit Bauer & Walter Müller

Production of the Harmonised European Time Use Database

Iiris Niemi; Statistics Finland, Finland

Hannu Pääkkönen; Statistics Finland, Finland

Pirjo Hyttiäinen; Statistics Finland, Finland

Paavo Väisänen; Statistics Finland, Finland

European Time Use Surveys were harmonised by Eurostat in 2000. Over 20 countries carried out a survey using an input harmonised diary and output harmonised background variables. In order to enlarge the number of potential users of the data a HETUS database is in the process of being compiled by Statistics Finland and Statistics Sweden. The final database will include time use micro data from 15 European countries, for which data are now completed for eight: Finland, France, Germany, Italy, Norway, Spain, Sweden and the United Kingdom. By the end of 2007, the following countries will be added to the database: Bulgaria, Estonia, Hungary, Latvia, Lithuania, Poland and Slovenia. The responsibility of Statistics Finland is to harmonise the data files while Statistics Sweden bears the responsibility for developing the web-based tabulation system and for the maintenance of the database. The action has received funding from the European Community.

For each 10 minute slot the diary database includes data on the main and secondary activity, location and “with whom” variables. Three-digit main activity codes were converted to the aggregate classification containing 49 comparable time use categories. Several background variables were included, covering basic individual and household indicators. Meta information concerning the applied statistical methods was appended to the database to help in understanding the structure of the sample design, the estimation procedures and the data collection.

Harmonisation problems have been observed during the process. Even minor national deviations from the detailed harmonised guidelines influence the validity of the comparative results.

This presentation is a demo of the web-based tabulation system.

Harmonization of the Collaborative Psychiatric Epidemiology Surveys (CPES).

Sue Ellen Hansen; University of Michigan, United States

The three component studies of the National Institute of Mental Health Collaborative Psychiatric Epidemiology Surveys (CPES) – the National Comorbidity Survey Replication (NCS-R), the National Latino and Asian American Study (NLAAS), and the National Study of American Lives (NSAL) – have both overlapping and unique content, similarities and differences in variable naming conventions, and differences in question wording and response options on some common measures. The Survey Research Center (SRC) and the Interuniversity Consortium for Political and Social Research (ICPSR) at the Institute for Social Research (ISR) collaborated on development of interactive Web-based documentation to facilitate understanding and analysis of the merged CPES public use data. This presentation describes challenges faced in harmonizing the three datasets, clarifying similarities and differences, and developing the Web-based documentation. It also describes the processes and tools developed to assist in harmonization, which we view as first steps in the development of best practices for harmonization of datasets at ISR.

1.4 Data Collection in Cross-cultural and Cross-national Surveys

Session coordinated by:

Annelies Blom; Centre for Survey Research and Methodology (GESIS-ZUMA), Germany

Achim Koch; Centre for Survey Research and Methodology (GESIS-ZUMA), Germany

Beth Ellen Pennell; University of Michigan, United States

Presentations

1. *Longitudinal Data Collection in Continental Europe: Experiences from the 2006 Survey of Health, Ageing and Retirement in Europe (SHARE)* – A. Börsch-Supan (Germany), K. Hank (Germany), H. Jürges (Germany), M. Schröder (Germany)
2. *The Use of Incentives on the European Social Survey* – M. Johnson (United Kingdom)
3. *Some Questions Raised While Piloting a Survey on International Migrations Conducted in Senegal and Europe* – M. Quaglia (France)
4. *Data Collection Challenges in Australian Remote Indigenous Communities* – T. Lalor (Australia), T. Wardrop (Australia)
5. *Sampling and Data Collection in Tribal Research - Issues and Strategies Undertaken* – O. Sajitha (India)

Presentations in the session "Data Collection in Cross-cultural and Cross-national Surveys"

Longitudinal Data Collection in Continental Europe: Experiences from the 2006 Survey of Health, Ageing and Retirement in Europe (SHARE)

Axel Börsch-Supan; Mannheim Research Institute for the Economics of Aging, Germany

Karsten Hank; Mannheim Research Institute for the Economics of Aging, Germany

Hendrik Jürges; Mannheim Research Institute for the Economics of Aging, Germany

Mathis Schröder; University of Mannheim, Germany

The Survey of Health, Ageing and Retirement in Europe (SHARE) collects data on the health, social, and economic situation of the population aged 50 or older in 15 European countries (including Israel). A baseline wave of data collection was conducted in 11 countries in 2004. Subsequently, SHARE expanded geographically and started a second – longitudinal – wave of data collection in the autumn of 2006, which will be completed in spring 2007.

Our presentation starts with an overview of the preparatory measures taken to transform the cross-sectional SHARE into a longitudinal enterprise: development of a longitudinal questionnaire (including preloading issues, exit interview, etc.), sample management system (including handling of split households, refresher sample, etc.), training concept, and so forth. Then we describe the experiences made during the survey's field period, focusing on the monitoring of progress and data quality in all countries. We conclude with preliminary results on response rates (including the SHARE drop-off questionnaire) and attrition.

The Use of Incentives on the European Social Survey

Mark Johnson ; National Centre for Social Research - NatCen, United Kingdom

The ESS is the most methodically rigorous cross-national social survey in the world. One of the cornerstones of this rigour is its emphasis on a minimum target response rate of 70%. However, in many countries, Britain included, such a response rate is becoming increasingly hard to achieve. This is true of surveys on areas of enquiry not prone to poor response rates, such as health and children, and is even more pronounced on surveys that look at the kinds of issues ESS is devoted to, namely social and political attitudes. In each of the last three years the British Social Attitudes survey, which covers a similar plethora of topics to the ESS, has achieved response rates below 60%. ESS in Britain in Rounds 1 and 2 got response rates of 55% and 51% respectively. In light of this, ESRC funded an incentive experiment on Round 3 that would provide researchers with robust empirical evidence as to the most appropriate use of incentives for any future rounds of ESS, not to mention other social surveys. There were three incentive groups:

1. book of stamps; conditional £10 voucher
2. book of stamps; no voucher
3. no stamps; conditional £10 cash

A particular innovation was a further experiment at the reissue stage of fieldwork. Half of the addresses in each incentive group had £5 added to the value of the incentive they were due to receive once they had completed the interview. The other half of addresses received the same incentive as they were initially allocated. Therefore, six incentive groups existed for reissues.

In this paper we will show the effect of different incentive strategies on response rates. However, it is increasingly being recognised that response rates are not necessarily the best benchmark for the quality of social surveys. Instead, a growing emphasis is being placed on non-response bias. Therefore, we will also look at the representation of different population subgroups within the different incentive

groups, aiming to establish whether particular incentive strategies are better at getting traditionally under-represented groups to participate. Further, we will look at another dimension of data quality, namely missing data, to try to establish whether particular incentives minimise its prevalence.

Some Questions Raised While Piloting a Survey on International Migrations Conducted in Senegal and Europe

Martine Quaglia; Institut National d'Etudes Démographiques, France

This paper aims to discuss some of the questions that arose within the first two test periods of the MAFE (Migrations between Africa and Europe) survey which is currently conducted by IPDSR (Institut de Population/Développement et Santé de la Reproduction, Dakar) and INED (Institut National d'Etudes Démographiques, Paris) in Senegal and in France, Italy and Spain. One of the main 'MAFE' project purpose is to participate in the current discussions on the impact of international migrations by producing actualised and objective quantitative data on the circulation of migrants, as well as on the impact of migration on the country of origin.

For this survey, two different questionnaires have been designed to be addressed to households in Senegal on one hand, to individuals in Senegal and France on the other. The interviewers training, the control on questionnaires which was done while the data was collected as well as the debriefings organized at the end of the data collection period, have brought out many questions related to the question of context. Political and cultural, national, context will be explored and discussed as being definitely very significant for a survey dealing with interactions between international migrations and development, and in which the question of comparison with other international surveys results is also of major importance.

Through a few examples of differences due to contextual evidences that have to be taken into account, and of questions that occurred while testing the procedure and the questionnaire, we will discuss some of the choices that have been made or still have to be done before the pilot, planned for mid July.

How to use existing 'know-how' from the national institutions, in terms of questionnaire design and implementation, interviewers training or data collection, and adapt ourselves and our own skills, keeping in mind the peculiar contexts in which the survey is being done will be the major subject of this discussion.

Key words : cross-cultural survey methodology, cross-cultural interviewer training, cross-cultural questionnaire design

Data Collection Challenges in Australian Remote Indigenous Communities

Thérèse Lalor; Australian Bureau of Statistics, Australia

Tim Wardrop; Australian Bureau of Statistics, Australia

Australia faces an interesting situation with regards to collecting data in remote Indigenous communities. There are many areas in Australia that are very sparsely populated. These areas are where a relatively high proportion of our Indigenous population lives. Collecting data from this population can be very challenging. There is no single Indigenous culture in Australia. In fact, there are over 600 unique Indigenous "language groups". Each community has its own ways of doing things (even those that are in close proximity to each other), and as such there is no prescriptive set of rules for interacting with Indigenous communities. Currently, the ABS conducts surveys in remote Indigenous communities on social issues such as health and housing, and is currently developing a collection on children. The cost and difficulty in collecting data from these communities often prohibits these populations from being included in other ABS population surveys.

The difficulties of collecting data in these communities include:

- Interviewers can not arrive at a community and expect to conduct surveys as they would in other areas of Australia (e.g. metropolitan Sydney). In remote Indigenous communities, cultural rules

and protocols must be observed. Facilitators (who are usually members of the community) are used to help ABS interviewers get access to the people they need to, provide assistance with interviews, and provide information about the culture of the community. The presence of the facilitator during interviews can introduce a bias to the answers that respondents provide.

- The low levels of literacy in some communities have an impact on the way data is collected. Using self enumerated forms is not feasible in these communities which makes collecting data on sensitive issues such as substance use difficult. This introduces non sampling error such as social desirability into the answers given.
- The concepts that underlie many ABS surveys are not understood in the same way in these communities. Data is generally collected in terms of 'households' or 'dwellings'. These terms do not mean the same thing to people in remote Indigenous communities, where an attachment to a dwelling and assumptions about resource sharing within households operate differently to the wider Australian population.
- Differences in language use, literacy, and exposure to formal education means that collection instruments used in these areas regularly utilise questions which are different to those used in the wider Australian population to try to capture the same or similar variables. A significant challenge in these collections is to maintain comparability of output while using different input items.

In recent years, strategies to overcome some of these issues have been developed. The most notable was the development of the cultural awareness package that was used during our recent Census Indigenous Post Enumeration Survey.

Sampling and Data Collection in Tribal Research - Issues and Strategies Undertaken

OG Sajitha ; Gujarat Institute of Developmental Research (GIDR), India

The cultural sensitivity of the population to be studied is an important aspect to be considered during sample design and data collection. It affects not only the data collection process but also the quality of the data collected itself. But how far it is possible to incorporate the cultural sensitivity of the respondents in various stages of research design is yet to be evolved. In the present paper an attempt is done to depict the issues faced during sampling and data collection for a research "Reproductive Health of Tribal women in Kerala, Prospects and Challenges". The paper describes the problems expected and encountered together with the measures taken while studying three tribal groups with heterogeneous behaviours even though located within the state boundary of Kerala (A state in India). The three tribal groups studied are Kanikkaran - a group having regular contact and interaction with non tribes, Adiya - a backward group and Kattunayakan- a primitive tribal group. The issues and strategies undertaken in the study would help to strengthen the trials of researchers and statisticians in the venture of evolving a feasible incorporation of cultural sensitivity in sample design and data collection among heterogeneous groups.

1.5 Sampling for Cross-national and Multi-frame Surveys

Session coordinated by:

Sabine Häder; Centre for Survey Research and Methodology (GESIS-ZUMA), Germany

Siegfried Gabler; Centre for Survey Research and Methodology (GESIS-ZUMA), Germany

Seppo Laaksonen; University of Helsinki, Finland

Peter Lynn; University of Essex, United Kingdom

Presentations

1. *Combining Landline Phone Samples and Mobile Phone Samples: a Dual Frame Approach* – S. Gabler (Germany)
2. *Clustering and Stratifying* – S. Laaksonen (Finland)
3. *Estimating Design Effects in the European Social Survey - Taking a Closer Look on Some Estimators of the Intra-class Correlation Coefficient* – M. Ganninger (Germany)
4. *Selecting of Non-overlapping Samples Using Probability Proportional to Size Systematic Sampling* – M. Kustura (Croatia)

Presentations in the session "Sampling for Cross-national and Multi-frame Surveys"

Combining Landline Phone Samples and Mobile Phone Samples: a Dual Frame Approach

Siegfried Gabler; Centre for Survey Research and Methodology (GESIS-ZUMA), Germany

The proportion of households exclusively using mobile phones has increased in Germany in the last years. Latest estimations indicate that 7- 8 percent of households have mobile telephones only. A way for future survey research could be to supplement landline phone samples by adding mobilephone samples. However, the question arises how to combine both samples. To solve this problem the German Research Foundation is funding a joint research project of the TU Dresden and the Centre for Survey Research and Methodology, Mannheim. In the paper we present results of some simulations that show the usefulness of the dual frame approach for combining both samples.

Clustering and Stratifying

Seppo Laaksonen; University of Helsinki, Finland

There are different strategies in human surveys for sampling designing as far as stratification and the choice of primary sampling units (psu) are concerned. A psu can be either directly a survey unit or it is a cluster in which case a survey unit is found within a psu. In the latter case, there are in principle a big number of alternatives to choose a psu including its size. I do not know which type of a pattern of psu's could be ideal although they are often small areas (such as census districts) of a country. But what is small and how different such psu's could be ideally? This has not been much discussed in the survey literature. Sometimes psu sizes vary substantially, as my experience from the European Social Survey (ESS) shows, even so much that some psu's are really small areas with some hundreds of survey units, whereas other areas have some hundred thousands units. This means that the latter areas could be rather handled as strata. Consequently, there can be some confusion with these basic concepts, that is, what is psu and what is stratum. This paper makes attempts to clarify this question while it presents an illustrative empirical exercise.

Estimating Design Effects in the European Social Survey - Taking a Closer Look on Some Estimators of the Intra-class Correlation Coefficient

Matthias Ganninger; Centre for Survey Research and Methodology (GESIS-ZUMA), Germany

Most statistical estimators yield unbiased and consistent estimates of a population parameter if calculated on the basis of a simple random sample (srs). If the same estimators are calculated on the basis of complex sample survey data, their variances are usually underestimated. In such a case, the design effect, $deff$, gives the factor of the magnitude of underestimation. Given a complex sample design, the design effect can be decomposed into the design effect due to unequal selection probabilities ($deffp$) and the design effect due to clustering ($deffclu$) (Gabler et al. 1999). The latter quantity, $deffclu$, can be estimated for the usual estimator of the mean, \bar{y} , using a formula that incorporates the mean cluster size, \bar{m} , and the intra-class correlation coefficient, ρ . Estimation and interpretation of this coefficient is a heavily debated topic (Choi 1987, Clemmer & Kalsbeek 1984, Irwin 1946, Killip et al. 2004, McGraw & Wong 1996, Srivastava & Keen 1988, Thomas et al. 1983, Walsh 1947). Numerous estimators have been proposed, each especially appropriate for certain kinds of data. Unfortunately, no attempt has yet been made to give a comprehensive overview of the most important estimators with respect to typical sample survey conditions. What the simulation studies, however, show is that the ANOVA estimator is usually an unbiased, efficient and consistent estimator of ρ (Mak 1988, Ridout et al. 1999, Paul et al. 2003). These empirical findings, together with its appealing intuitive interpretation and its computational simplicity are the reasons why for the estimation of design effects in the ESS, we use the ANOVA estimator. It is, however, worthwhile taking a closer look at other estimators too and to reveal

the operating conditions under which their potentials and weak points become most obvious. In this paper, I will make two central points: first, I will show the results of a simulation study that compares the behaviour of different estimators under varying, yet typical sample survey conditions (i.e. different (and unequal) cluster sizes, variable scaling (continuous, discrete, binary), magnitude of homogeneity in the population, etc.). Second, I derive recommendations for the survey practitioner based on examples using ESS data.

Selecting of Non-overlapping Samples Using Probability Proportional to Size Systematic Sampling

Marinela Kustura; Statistics Croatia, Croatia

Here we demonstrate how we selected non-overlapping samples of area units in the first stage of selection for Croatian Labour Force Survey (LFS) 2007, using probability proportional to size (PPS) systematic sampling.

The EU Labour Force Surveys use rotating sampling designs, in order to provide estimates of changes in the characteristics of the labour force over time. In the rotating sampling design, a sample consists of number of sub-samples or rotation groups. Each survey round (i.e. quarter in the LFS), some rotation groups are replaced, according to a pre-defined rotation pattern.

According to the COUNCIL REGULATION (EC) No 577/98 of 9 March 1998 on the organisation of a labour force sample survey in the Community, the survey is a continuous survey providing quarterly and annual results; the reference weeks are spread uniformly throughout the whole year; the reference quarters and years are respectively groups of 13 or 52 consecutive weeks.

From 2007, the Croatian LFS uses the rotation pattern 2-(2)-2. This means that in 6 successive quarters each rotation group is participating in the survey in the first two quarters and in the fifth and sixth quarter, but not in the third and fourth quarter.

The sample for the Croatian LFS is selected in two stages. In the first stage, a sample of area units is selected, and in the second stage a sample of dwellings is selected within the selected areas.

In order to accommodate a 2-(2)-2 rotation pattern and uniform spread of the sample throughout 52 weeks of the year, for LFS 2007 nine rotation groups had to be selected. This was done by selecting nine rotation groups of area units, and each rotation group was divided into 13 weekly sub-samples.

We applied PPS systematic sampling method for sampling area units of varying size, and systematic sampling method with equal probability for sub-sampling area units for weekly samples.

Because of PPS systematic sampling, we had to consider how to select nine non-overlapping samples of area units. It was done by writing a SAS program for PPS systematic sampling using a fractional interval method to provide a sample of specified size, and varying a random starts for different samples.

1.6 Methodological Issues in Multilevel Analysis for Cross-national Research

Session coordinated by:

Bart Meuleman; University of Leuven, Belgium

Presentations

1. *The Small N-problem in Cross-national Survey Research: Accuracy Assessment of Multi-level SEM by Means of a Monte Carlo Study* – B. Meuleman (Belgium)
2. *Multilevel Structural Equation Modeling with Small Group Sample Sizes: Evaluating Model Stability and Accuracy* – T. Reeskens (Belgium), B. Meuleman (Belgium)
3. *A Cross-national Analysis of the Impact of Relative Income Position on Consumers' Offending Practices* – C. Abreu Lopes (United Kingdom)
4. *Modeling Educational Success As a Function of Both Class and Student/pupil Characteristics: a Hierarchical Approach.* – M. Jelonek (Poland)

Presentations in the session "Methodological Issues in Multilevel Analysis for Cross-national Research"

The Small N-problem in Cross-national Survey Research: Accuracy Assessment of Multilevel SEM by Means of a Monte Carlo Study

Bart Meuleman; University of Leuven, Belgium

Thanks to the increasing availability of international survey data (e.g. the European Values Study and the European Social Survey), there exists growing scientific interest for cross-national comparisons of values, attitudes and opinions. Various scholars have used international surveys to link individual characteristics to aspects of the national context. Often, multilevel techniques are applied to explain individual-level variables by means of country-level features.

However, the application of multilevel models in the field of cross-national research is far from unproblematic. Due to budget limitations, the number of participating countries does not exceed 25 for most international surveys. Consequently, the group level sample sizes are often substantially lower than what rules of thumb suggest (at least 50 or 100 units). On the other hand, cross-national surveys typically contain a large number of respondents per country (> 1000).

This paper summarizes the results of a Monte Carlo study that was carried out to assess the accuracy of multilevel modeling in the domain of cross-national research. More specifically, the study concentrates on a rather recent but very promising statistical tool, namely multilevel structural equation modeling (SEM). A multilevel SEM, in which a latent factor is explained by a within- and a between-level variable, is simulated. In order to reproduce realistic circumstances as much as possible, the situation of the ESS round 1 (2002-2003) -22 countries and over 40.000 respondents- is taken as a starting point. The size of the between-level variable effect and the intra-class correlations are manipulated. In order to test whether trade-off effects between individual sample size and group sample size are present, various numbers of countries and respondents per country are simulated. For all conditions, the parameter estimates and their respective standard errors for both the within- and between model are evaluated. Special attention is given to the power for detecting the effect of the between-level variable.

References

Hox, J.J. & Maas, J.C.M. (2001). The accuracy of multilevel structural equation modeling with pseudobalanced groups and small samples. *Structural equation modeling*, 8(2), 157-174.

Multilevel Structural Equation Modeling with Small Group Sample Sizes: Evaluating Model Stability and Accuracy

Tim Reeskens; University of Leuven, Belgium

Bart Meuleman; University of Leuven, Belgium

Due to recent software developments, it has become relatively easy to estimate complex multilevel structural equation models. According to Rabe-Hesketh, Skrondal and Pickles (2004), "multilevel structural equation modeling is required when the units of observation form a hierarchy of nested clusters and some variables of interest cannot be measured directly but are measured by a set of items or fallible instruments." Clearly, these advanced statistical analysis techniques have several desirable characteristics for cross-national attitude research, as they are fit to analyze nested data structures (individuals are clustered within countries) and to model latent variables simultaneously.

Together with these recent innovations in multilevel structural equation models, important methodological issues have come to the fore. One of these issues is sample size. In conventional multilevel analysis, the '30/30 rule' by Kreft (1996) is often used as a rule of thumb. According to this guideline, at least 30 groups with a minimum of 30 individuals per group are necessary to obtain reliable parameter estimates (Hox, 2002). Yet, when there is a special interest in cross-level interaction terms or random

effects, it is recommended to have a larger sample at the second level, for instance about 100 groups with 10 individuals per group (Hox, 2002). Because multilevel structural equation models often have a high degree of complexity, sample sizes required for accurate estimation might be even larger.

In cross-national surveys, the number of participating countries usually does not exceed 25. Therefore, the application of multilevel structural equation models in the domain of cross-national research can pose estimation problems. However, it is not known to what extent the accuracy and stability of the models is affected by the small N-problem.

In this paper, the stability of multilevel structural equation models with decreasing sample sizes is tested. We use data of the 2006 Belgian Youth Survey (Hooghe et al., 2006). In this survey, data on 6000 16-year olds clustered within 112 schools is available. In a first step, a multilevel SEM (with as dependent variable ethnocentrism) is estimated for the total sample. Because this total sample size is adequate according to common guidelines, we expect the parameter estimates to be reliable. In subsequent steps, the group sample sizes are decreased by means of subsampling (100, 80, 60, 40 and 20 schools). This procedure makes it possible to decide how many groups at the between-level are necessary to obtain stable and accurate results.

References

- Hooghe, M., Claes, E., Dejaeghere, E., & Quintelier, E. (2006). Belgian Youth Survey. Data Set. Leuven: KU Leuven.
- Hox, J. (2002). *Multilevel Analysis: Techniques and Applications*. Mahwah: Erlbaum.
- Kreft, J. J. (1996). *Are Multilevel Techniques Necessary? An Overview, including Simulation Studies*. Unpublished Paper. Los Angeles: California State University.
- Rabe-Hesketh, S., Skrondal, A., & Pickles, A. (2004). Generalized Multilevel Structural Equation Modeling. *Psychometrika*, 69(2), pp. 167-190.

A Cross-national Analysis of the Impact of Relative Income Position on Consumers' Offending Practices

Cláudia Abreu Lopes; London School of Economics, United Kingdom

This paper aims to examine the effect of individual position in income distribution on consumers' dishonest behaviour from a cross-national perspective. These unscrupulous, but not necessarily illegal, practices have burgeoned as a result of the transformations of European economy in the late twentieth century, specifically the transition to neo-liberal markets and emergence of consumer society (Karstedt & Farral, 2004). A "cornucopia of new criminal opportunities" has given room to actions such as paying cash-in-hand to circumvent taxes, making false insurance claims or over-claiming government benefits. Notwithstanding they are collectively regarded as morally dubious, they are part of people's experience. Often referred as "crime of everyday life", they are committed mostly in the marketplace by those who see themselves and are seen as respectable citizens (Gabor, 1994; Karstedt & Farral, 2006). Consumers' unethical practices could be merely seen as a form of expressing discontentment with business and market institutions whose main actors would be those who are at the centre of consumer societies, that is to say, members of middle classes. According to Karstedt & Farral (2004), they are indicators of the moral stage of society and therefore a worthwhile instrument for social and political analysis.

The importance of relative income position on individual attitudes and behaviour has been stressed in social sciences particularly in studies related to happiness and well-being (Ferrer-i-Carbonel, 2005; O'Connell, 2004). Typically, this kind of studies employs individual data from one single country. As an exception, a study by Fisher and Togler (2006), using data from 1998 ISSP survey, presented strong evidence in favour of an association between relative income position and social capital variables (social trust, trust in institutions, social norms and networks). Following this idea, the present paper assumes that the consumers' level of engagement on unethical practices depends on the distance between their own and the reference group's income.

The empirical study consists on a cross-national analysis of consumers' offending practices using data from ESS (Round 2) that focuses on economic morality, financial and socio-economical variables. Various measures of relative income positions are discussed and contrasted. A measure based on individual income deviation from the average of socio-structural group of which individual belongs is proposed.

A multi-level random coefficient model (taking country as clustering variable) is presented in order to shed light on the relationship between relative income position and consumers' offending in a comparative perspective. The model also includes a set of country level covariates (Laeken indicators) selected from EU-SILC to better isolate partial effects. From a methodological point of view this paper contributes, on the one hand, to expand the scope of the test of the relative-income hypothesis from one country to a cross-national standpoint and, on the other hand, to reinforce the statistical analysis of cross-country surveys by integrating micro and macrolevel rigorous data.

References

- Ferrer-i-Carbonell, A. (2005). Income and well-being: An empirical analysis of the comparison income effect. *Journal of Public Economics*, 89, 997-1019.
- Fischer, J A. and Torgler, B. (2006). The effect of relative income position on social capital. *Economics Bulletin*, 26 (4), 1-20.
- Gabor, T. (1994). *Everybody does it! Crime by the public*. Toronto: University of Toronto Press.
- Karstedt, S. & Farral, S. (2004). The moral maze of the middle class: The predatory society and its emerging regulatory order. In H.-J. Albrecht, T. Serassis, & H. Kania, *Images of crime II* (pp. 65-94). Freiburg: Edition Iuscrim.
- Karstedt, S. & Farral, S. (2006). The moral economy of everyday crime: Markets, consumers and citizens. *British Journal of Criminology*, 46 (6), 1011-1036.
- Kosicki, G. (1987) A test of the relative income hypothesis. *Southern Economic Journal*, 54, 422-434.
- O'Connell, M. (2004) Fairly satisfied: Economic equality, wealth and life satisfaction. *Journal of Economic Psychology*, 25, 297-305.

Modeling Educational Success As a Function of Both Class and Student/pupil Characteristics: a Hierarchical Approach.

Magdalena Jelonek; Jagiellonian University, Poland

The aim of this paper is to present the technique overcoming problems connecting to modeling social structures which are typically multilevel – Hierarchical Linear Modeling. In Social Sciences it is used for the analysis of a variety of problems in educational, organizational research, but also market research. It is especially of a great interest in handling multilevel, cross-national data.

HLM provides a conceptual and statistical mechanism for investigating and drawing conclusions regarding the influence of phenomena at different levels of analysis. Typically these structures are naturally occurring ones: workers populations are characterized by individuals nested within department, student/pupils nested within classes.

This paper illustrates how a hierarchical linear model can be used by researchers and practitioners to estimate educational success as a function of both class and student/pupil characteristics.

The presentation, based on data from regional examination commission (Cracow) focuses on comparing two models: hierarchical and non – hierarchical one, presenting some challenges researchers faced while applying this technique.

1.7 Measurement Quality of Concepts in the ESS

Session coordinated by:

Willem E. Saris; ESADE, Universitat Ramon Llull, Spain

Presentations

1. *Measurement Requirements for Comparative Research* – W. Saris (Spain)
2. *A Cross-national Comparison of the Trust in People Scale* – W. van der Veld (Netherlands)
3. *The Measurement of Inter-personal and Social Participation in the ESS* – L. Guillén (Spain)
4. *The Comparability of the Human Values Measurements in the Second Round of the European Social Survey* – E. Davidov (Germany)
5. *The Quality of the Measurement of Media Use in the ESS* – L. Coromina Soler (Spain), W. Saris (Spain), D. Oberski (Spain)
6. *The Measurement of Political Trust in the ESS* – D. Oberski (Spain)

Presentations in the session "Measurement Quality of Concepts in the ESS"

Measurement Requirements for Comparative Research

Willem E. Saris; ESADE, Universitat Ramon Llull, Spain

The measurement requirements for comparative research specified in the literature for concepts with reflective indicators are: configural invariance, metric invariance and scalar invariance. Configural invariance is obtained if the same standard factor analysis model should hold for all different groups. One has to add to this requirement the equality of the loadings to speak of metric invariance and to obtain scalar invariance the intercepts should also be equal in the different groups.

In this presentation we argue that these requirements are too strict. We will specify a response model that makes a distinction between a measurement part of the response process and the cognitive part. We will argue that it can happen that the above specified requirements are not satisfied because of differences in the measurement process or in the cognitive part or in both. It is essential that the cognitive part is the same across groups because otherwise people have different ideas about the concepts of interest. Differences in the measurement part are less fundamental. One can estimate the differences in this measurement process separately and correct for these differences.

As a consequence we suggest that the above mentioned requirements for comparative research should hold after correction for measurement errors.

A Cross-national Comparison of the Trust in People Scale

William van der Veld; University of Amsterdam, Netherlands

Social trust is a key expression of civic virtues which are necessary for an effective and stable democratic government. Lower levels of social trust are assumed to be related to lower stability and effectiveness in democratic government. This hypothesis can only be fruitfully studied across nations (and/or time). Cross-national comparative research requires that the measurement instruments are equivalent across the nations. The most often used measure of social trust is a single item measure that was developed by Elisabeth Noelle-Neumann in 1948 in Germany. The single item measure is part of the European/World Values Study and is therefore often used in cross-national comparative studies. This despite the fact that it is not known whether this measure is equivalent across nations. Equivalence in the single item case means that the quality of the measure (Van der Veld, 2006) is the same across countries. If there are differences in the quality, one cannot compare across countries in a straightforward way without correction for the lack of quality. In this study we will test whether previous studies were correct in assuming that the quality is the same across countries. This assumption is tested by estimating the quality of this measure with the split ballot MTMM (Saris, Satorra, and Coenders, 2004) studies carried out in round 1 of the European Social Survey (ESS).

Besides the single item measure there are also various multi-item scales that measure social trust. One that is often used is the 3-item trust-in-people scale (Survey Research Centre, 1969) which is an adapted version of Rosenberg's 5-item faith-in-people scale (1957). We will also test whether the trust-in-people scale is equivalent across nations. This social trust scale is also part of the core of the ESS. For the test of the equivalence of social trust across the participating countries in the ESS we use a procedure that is similar to the procedure suggested by Steenkamp and Baumgartner (1998). This procedure involves tests of invariance (configural, metric, scalar, and error). Invariance refers to whether or not under different conditions of observing and studying phenomena measurement operations yield the same attribute (Horn & McArdle, 1992). In the test of metric invariance they (Steenkamp & Baumgartner) suggest to constrain the factor loadings of each indicator to be equal across nations. This constraint is an ambiguous constraint, because it assumes that the product of the quality and the consistency is equal across countries. Where the consistency is an indication of amount of uniqueness of each indicator. For example each of the indicators (trust, fair, helpful) of the trust-in-people scale measure

social trust and something else, a unique component that is substantive by nature. Because cross-national invariance involves the measurement of the same attribute across nations, it is not important to assume equal quality across nations. Therefore in the test we perform we will correct all indicators for lack of quality by first estimating the quality with the split ballot MTMM studies carried out in round 1 of the European Social Survey (ESS).

The Measurement of Inter-personal and Social Participation in the ESS

Laura Guillén; ESADE, Universitat Ramon Llull, Spain

The notion of participation has been studied extensively from different perspectives, but the quality of its measures has not received much attention. In this paper we propose an approach to measure participation in the European Social Survey (ESS) focusing in the integration of the individual into broader economic and social networks (Du Toit, 2004).

The proposed model will be tested and the quality of the ESS measures with respect to reliability, validity and composite scores will be studied. In addition, Multiple-group analysis for different countries will be used to assure comparability between countries.

Finally, we will check external validity of these measures by estimating the relationship between the deprived conditions in the ESS countries (including financial poverty, health, education and employment) and participation.

The Comparability of the Human Values Measurements in the Second Round of the European Social Survey

Eldad Davidov; University of Köln, Germany

Human values have played an increasingly important role in social science research in the last decade. There are two main reasons for this development. First, Shalom Schwartz (1992) introduced a new value theory, which proposes 10 universal values and four value dimensions. Furthermore, he developed two instruments to measure them which were successfully tested in many countries. Secondly, the European Social Survey (ESS) decided to adopt one of these instruments and include a shortened version of it in its core questionnaire. This instrument was implemented in 20 countries in the first round of the ESS (2002/3) and in 25 countries in the second round (2004/5), and it is going to be included in the future as well. Researchers may use it to study values and their associations with other theoretical constructs of interest.

A central question when applying this instrument in cross-country research is whether and to what extent it is equivalent across countries. Measurement (metric) invariance, is a necessary condition to insure equivalence of the meaning of the constructs. It is a precondition for comparing associations between values and other constructs across countries. Scalar invariance should also be assessed, as it is a precondition for comparing means across countries. These tests were conducted by Davidov, Schmidt and Schwartz (in press) using the first round of the ESS. However, values have not been subjected to such tests in the second round.

In this study we use multi-group confirmatory factor analysis (MGCFA) of data from the second ESS round, to assess configural, measurement (metric) and scalar invariance of the values across 25 countries. The MGCFA did not support configural invariance of a model with 25 countries. After reducing the number of countries to 14, the analyses supported configural and metric invariance of a model with 7 values. High construct intercorrelations necessitated unifying 3 pairs of values from the theory. However, these values are adjacent in the value circle of the Schwartz theory, and share a broad basic motivational goal. Four items intended to measure particular value constructs also had significant, secondary loadings on other value constructs. The analysis did not support scalar invariance. Researchers may now use the ESS value scale of the second round to confidently carry out cross-country comparisons of the relationships between values and other variables such as socio-demographic characteristics, attitudes and behaviour across the 14 countries. However, comparison of mean values still remains problematic. It may still be possible to find pairs or triads of countries where scalar invariance holds.

Key words: European Social Survey; configural, metric (measurement) and scalar invariance; partial invariance; human values.

The Quality of the Measurement of Media Use in the ESS

Lluís Coromina Soler; ESADE, Universitat Ramon Llull, Spain

Willem E. Saris; ESADE, Universitat Ramon Llull, Spain

Daniel Oberski; ESADE, Universitat Ramon Llull, Spain

Many studies in different fields study media use and purposes of the time spent on media. Most of these studies are not concerned in the quality of the measures but just on the use of the information. The aim of this paper is to study the quality of the media use measurement in the European Social Survey (ESS) for cross cultural research. The ESS measure for media use is based on a proposal of Ken Newton. The total time and time spent on political issues on television, radio and newspaper has been asked. This is first asked in general with respect to program concerning political issues.

In the ESS these questions have been asked in different ways which allows a Multitrait Multimethod (MTMM) analysis for each country to determine the reliability and validity of these questions in all countries. The information from the different MTMM designs will be used for the Multiple-group analysis on different countries. This multiple-group analysis will be used to test for equivalence (metric invariance and scalar invariance). This analysis enables us to indicate which comparisons between which countries can be made. In a next step composite scores will be calculated for media use and the quality of these indices will be determined.

Finally, we will check the external validity of these measures by estimating the relationships with variables that are expected to be correlated with media use and political issues on media. In this context we will use variables available in the ESS questionnaire such as political interest, political knowledge, personal discussion about political issues and political participation.

The Measurement of Political Trust in the ESS

Daniel Oberski; ESADE, Universitat Ramon Llull, Spain

1.8 Developing Attitudinal Indicators

Session coordinated by:

Roger Jowell; City University, United Kingdom

Gillian Eva; City University, United Kingdom

Presentations

1. *Is There a Need for Attitudinal Social Indicators?* – R. Jowell (United Kingdom), G. Eva (United Kingdom)
2. *The Inter-personal Determinants of Regional Well-being in Europe* – L. Corrado (United Kingdom), A. Aslam (United Kingdom)
3. *Subjective Indicators of Well-being: Where Happiness Measures Fall Short* – L. Hyman (United Kingdom)
4. *Coping with Complexity: the Role of Composite Indicators* – M. Nardo (Italy), G. Munda (Italy), A. Saltelli (Italy)
5. *What About Index Construction?* – J. Boelhouwer (Netherlands)
6. *Composite Indicators: Science or Artifacts?* – M. Saisana (Italy), F. Cartwright (Canada)

Presentations in the session “Developing Attitudinal Indicators”

Is There a Need for Attitudinal Social Indicators?

Roger Jowell; City University, United Kingdom

Gillian Eva; City University, United Kingdom

Most current official indicators of the relative success and failure of nations - such as GDP, RPI, growth rate, borrowing requirement, and so on – are economic in origin. These indicators are clearly important, since the strength and direction of the economy have a large bearing on all aspects of life, including social provision. Even so, economic factors are by no means the only indicators of national success or failure. A strong and thriving economy is no guarantee of a sustainable high quality of life for many citizens, just as a relatively fragile economy is not necessarily associated with a low quality of life. If we wish to measure social or psychological aspects of quality of life, then we need to measure them directly.

Early work concentrated heavily on trying to find a single overall psychological measure of ‘happiness’ or ‘life satisfaction’ and was greeted with deep suspicion by most policy makers. As a result, little progress was made in promoting social indicators as official measures. The situation has now changed dramatically. Stimulated largely by the OECD and the EU, but also by several national governments, a great deal of work is now underway to develop appropriate social indicators – mostly, but not entirely, composite indicators, such as the Genuine Progress Indicator and the Human Development Index. Following work by Sir Tony Atkinson for the EC in 2000 (see Atkinson et al, 2002), 18 indicators of social inclusion/exclusion were adopted by the EU and are now routinely collected by member states. And a great deal of complementary work is now taking place to find other composite social indicators.

However, almost all of the social indicators being considered are factual or behavioural indicators. For instance, just one of the eighteen Atkinson-inspired indicators (a well-established measure of subjective health state) is attitudinal. Yet just as a society’s success should not be measured by economic factors alone, so – we contend – should it not be measured by factual and behavioural factors alone. Since the aim of the EU is to improve the quality of life of its citizens, we should try to measure quality of life at least partly through the eyes of its citizens.

This paper describes a programme of work aimed at compiling a long-list of candidate attitudinal indicators on different aspects of life that - individually or collectively – might become ‘official’ social indicators either in conjunction with or in addition to behavioural or factual social indicators. It considers the arguments for some of these attitudinal indicators and what they might add to the analysis of societal progress.

The Inter-personal Determinants of Regional Well-being in Europe

Luisa Corrado; University of Cambridge, United Kingdom

Aqib Aslam; University of Cambridge, United Kingdom

There is a strong need to supplement the existing, predominantly economic, analysis at the European regional level with an analysis of social well-being.

We seek to test the determinants of well-being, defined as short-term ‘happiness’ and longer-term ‘life satisfaction’. We also wish to test whether regional factors exert a stronger influence on individual well-being relative to national factors, and we use a theoretical model to justify such a premise.

In testing these assertions, our work proceeds to extend the measurement of well-being across European regions in several ways. Firstly, we draw on data from the European Social Survey and Eurostat to construct a hierarchical dataset grouped by area, and we assess the determinants of well-being using a multilevel modeling approach. Secondly, we account for the effects of social exchange within each group as well as intrinsic socio-demographic indicators and higher-level exogenous contextual factors.

Empirical findings support the idea that well-being is strongly dependent both on these general forms of social interactions and on more specific individual characteristics such as inter-personal feelings

and functioning, for example, the level of the respondents' trust in other people and their country's institutions. In addition, we find that there is some evidence of greater regional effects relative to national effects, but individual well-being continues to be affected most by micro-level phenomena.

Subjective Indicators of Well-being: Where Happiness Measures Fall Short

Laura Hyman; University of Surrey, United Kingdom

The last thirty years have seen a rapid proliferation of social scientific research into happiness and well-being. In particular, social indicators research has emphasised the importance of happiness measures at a national level, alongside more traditional economic or 'objective' measures like GDP-growth and per-capita income, for the monitoring of societal progress. Happiness measures are, at the very least, important measures to include alongside the more conventional economic indicators in national accounts, if we are to have a proper understanding of all dimensions of human well-being and progress. Interest in measuring and incorporating happiness into policy has been fuelled by well-publicised analyses of people's self-reported levels of happiness from population surveys like the European Social Survey. Such surveys ask generalised questions about happiness, along the lines of: "Taking all things together, how happy would you say you are, 00 being extremely unhappy and 10 being extremely happy?" Studies aggregating these measures to a national level contrast such measures with rates of GDP growth. They have found that if a country's level of GDP is high, it does not necessarily follow that levels of subjective well-being (such as feelings of happiness or satisfaction) will match this (Easterlin, 1974, Layard, 2005, Veenhoven, 1994).

This paper will firstly consider the assumption made in social indicators research that high levels of happiness represent societal progress. Why is this assumption made? How and why is happiness seen to be such an important 'goal' in contemporary western society? This paper will put forward some sociological arguments that help to answer these questions.

The paper will then go on to discuss how existing happiness research fails to acknowledge some of the shortfalls of happiness survey measures as social indicators. In particular, they fail to acknowledge the socio-cultural factors – such as social norms - that influence people's feelings of happiness. Concepts that are highlighted in the sociological literature will be presented and discussed.

A new methodology for measuring happiness, that primarily adopts a time-use approach, is proposed. This alternative to the survey measures would be able to capture and reflect socio-cultural influences on happiness, and would thus generate more reliable measures. Data obtained from this could also be of more use in informing policy decisions concerned with the advancement of societal progress.

Coping with Complexity: the Role of Composite Indicators

Michela Nardo; European Commission - Joint Research Centre, Italy

Giuseppe Munda; European Commission - Joint Research Centre, Italy

Andrea Saltelli; European Commission - Joint Research Centre, Italy

Composite indicators are increasingly used to summarize complex phenomena where a plethora of different dimensions have to be considered together. The main virtue of these measures is their usefulness in policy analysis in that they can illustrate multifaceted and sometimes elusive issues in wide ranging fields, e.g., environment, economy, society or technological development. Composites often seem easier to interpret than finding a common trend in many separate indicators and have proven useful in benchmarking country performance. However, composite indicators builders have to face a relevant degree of skepticism. They are accused of merging "apples and pears" when deriving a unique ranking from a set of indicators expressed in different measurement units. Furthermore the lack of transparency regarding, for example, the selection of the reference framework, the weighting scheme, the aggregation method, could produce misleading and simplistic analytical or policy conclusions. On the other hand a sound process of construction cannot remedy an inadequate framework or poor quality data. Therefore the

construction of a composite indicator requires a balance between different aspects, all equally important in defining the quality and finally the usefulness of the composite. This paper reviews the debate around composite indicators underlining the main steps to be followed in their construction. We do not claim the superiority of composite measures, but rather the necessity of soundness and transparency in their construction. This is particularly true in the field of social sciences where composite indicators are newcomers.

What About Index Construction?

Jeroen Boelhouwer; Social and Cultural Planning office (SCP), Netherlands

In the Netherlands, the Social and Cultural Planning Office uses a composite index to describe and monitor developments in the quality of life of the Dutch. Since 1974 this Life Situation Index (LSI) has been used to measure objective aspects of individual living conditions. Every two years the results are published in a report called “The social state of the Netherlands”.

In constructing the index, we take as our starting point the view that quality of life is a multi-dimensional concept. Therefore the LSI is composed of indicators from eight life domains: housing, health, leisure activity, consumer durables, sport activity, vacation, social participation and mobility. Education, income and paid employment are considered as “resources”: they are not part of the index itself, but are used to describe the backgrounds of the life situation.

The LSI is based on micro-data, stemming from survey research. As a result of this we are able to give a summarizing insight in the different domains of the life situation at the individual level. Moreover, we can break down the index for all the social groups one is interested in; not only on demographic grounds, but on economic or geographical grounds as well.

This paper / presentation, concentrates on three of the major issues we come across:

1. Why use a composite index? What are the advantages and disadvantages of using a composite index as opposed to single indicators?
2. The choice of domains and indicators. Does it really matter what domains are chosen? And does it matter which indicators are chosen within the domains?
3. How to combine the indicators. The LSI is nowadays constructed statistically, using nonlinear canonical correlation analysis (OVERALS in SPSS). But is this necessary to use such a rather difficult-to-explain technique: why can't we just add simply up the indicators? And why should weights be used at all?

Composite Indicators: Science or Artifacts?

Michaela Saisana; European Commission Joint Research Centre (JRC), Italy

Fernando Cartwright; Canadian Council on Learning (CCL), Canada

In recent years, numerous approaches to measure complex issues at the local and global scale have emerged in the form of composite indicators. A composite indicator aggregates multidimensional concepts that cannot be fully captured by any individual indicator alone. Some advantages of composite indicators include their ability to summarize multidimensional issues, generate ranked lists, track progress over time, facilitate public communication, gauge policy analysis, and promote accountability. Some disadvantages include their tendency to invite simplistic policy conclusions, disguise serious failings in certain dimensions and thereby hamper implementation of proper remedial action, and give rise to inappropriate policies if difficult-to-measure indicators of performance are ignored from the conceptual framework.

While the debate between advocates and detractors may never be settled, policy analysts will continue to develop composite indicators, and policymakers will continue to rely on them for decision-making. To maximize their utility and minimize their misuse, developers must be based on the best

available evidence, design the composite indicators with transparent structures, and assess the composite indicators using appropriate multivariate and sensitivity analyses. Sensitivity analysis acknowledges various methodological assumptions that are intrinsic to policy research and assesses whether results change substantially when those assumptions vary over a reasonable range of possibilities (Saltelli et al. 2007, Saisana et al. 2005).

Despite the importance of this methodological stage, critical issues of uncertainty, sensitivity and robustness receive minimal attention in the majority of composite indicators practices. Most index creators provide no evidence on, for instance, the choice of linear aggregation over multiplicative aggregation or the need for multiple levels of aggregation. The same is true for the logic behind indicator weights, indicator re-scaling, handling of missing values, and so forth. In particular, questions of indicator weighting and linear indicator aggregation require cautiousness due to the potential for indicator compensation. Overall, composite indicators science needs to consider and incorporate findings from disciplines such as econometrics and statistics to manage problems of uncertainty introduction, to strengthen index validity and to advance the methodological quality of those indices.

To discuss the points raised above, this presentation profiles selected indices from the fields of lifelong learning (Composite Learning Index, 2007), international trade (Trade and Development Index, 2005) and human health (Alcohol Policy Index, 2007). These indices were critically assessed and found to represent “good practices” in the composite indicator art-making.

The final goal is to propose an analytical framework for index evaluation that allows for the identification of scientifically sound and methodologically robust approaches based on recent reviews (Nardo et al. 2005, Gall M. 2007, Saisana and Tarantola 2002). Up to now, most indices remain in academic settings with limited applied exposure - except indices developed by large institutions such as the United Nations or the World Health Institute. In order to move beyond theoretical discussions of indices and their validity, it is necessary to implement the most promising indices as tools for detecting, monitoring and benchmarking performance in the relevant fields of interest. Only rigorous usage by the practitioner community will ultimately uncover reliability and feasibility flaws.

References

- Alcohol Policy Index (2007) published as Brand D.A., Saisana M., Rynn L.A., Pennoni F., Lowenfels A.B., 2007, Comparative Analysis of Alcohol Control Policies in 30 Countries, PLoS Medicine, 0759 April 2007, Volume 4, Issue 4, e151, p. 0752-0759, www.plosmedicine.org
- Composite Learning Index (2007) Official website, Canadian Council on Learning, www.ccl-cca.ca/cli.
- Gall, M. (2007) Indices of social vulnerability to natural hazards: A comparative evaluation, PhD dissertation, Department of Geography, University of South Carolina.
- Nardo, M., Saisana, M., Saltelli, A., Tarantola, S., Hoffman, A., Giovannini, E. (2005) Handbook on constructing composite indicators: methodology and users guide, 2005, OECD-JRC joint publication, OECD Statistics Working Paper, STD/DOC(2005)3, JT00188147, pp. 108.
- Saisana, M., Saltelli, A., Tarantola, S. (2005) Uncertainty and sensitivity analysis techniques as tools for the analysis and validation of composite indicators, *Journal of the Royal Statistical Society A*, 168(2), 307-323.
- Saisana, M., Tarantola, S. (2002) State-of-the-art Report on Current Methodologies and Practices for Composite Indicator Development, EUR Report 20408 EN, European Commission, JRC, Institute for the Protection and Security of the Citizen, Ispra, Italy, 72 pp.
- Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gabelli, D., Saisana, M., Tarantola, S. (2007) Global sensitivity analysis. Gauging the worth of scientific models. A textbook of methods to evidence how model-based inference depends upon model specifications and assumptions. Wiley (forthcoming)
- Trade and Development Index (2005) Developing countries in international trade. United Nations <http://www.unctad.org/>

2 Quality of Survey Questions

Sessions in Quality of Survey Questions

Software Presentations	42
Questionnaire Translation and Adaptation	44
Questionnaire Development and Testing	50
Communicating Survey Quality to Users	56
Interviewer and Respondent Behaviour in Survey Interviews	59
Augmenting Survey Data by Bio-indicators, Geo-data and Record-linkage	63
Improving the Validity of Psychological Questionnaires in the Context of Survey-based Research	65
Enhancing Survey Methodology with Qualitative Methods	70

2.1 Software Presentations

Daniel Oberski; ESADE, Universitat Ramon Llull, Spain

Presentations

1. *Sequence Viewer Demonstration* – Y. Ongena (Netherlands), W. Dijkstra (Netherlands)
2. *Survey Quality Predictor (SQP)* – D. Oberski (Spain), W. Saris (Spain), I. Gallhofer (Spain)

Presentations in the session "Software Presentations"

Sequence Viewer Demonstration

Yfke P. Ongena; Vrije Universiteit Amsterdam, Netherlands

Wil Dijkstra; Vrije Universiteit Amsterdam, Netherlands

In the demonstration we will give an overview of the wide variety of possibilities of the SEQUENCE VIEWER Program. This program was developed to manage, code and analyse sequential data originating from video, audio and/or text. First, we will give an overview of how the data in the program are organized. Such data can be behavioural data (e.g., verbal behaviour in survey interviews), but also other types of data (e.g., newspaper articles, job histories). Next, we will give an overview of the possibilities of the program with respect to data-manipulation (i.e., coding and auto-coding, labelling, fuzzy text search, editing text, and linking audio and video), and analysis. We will focus especially on the way recordings of verbal interaction of survey interviews can easily be transcribed, archived and coded in the SEQUENCE VIEWER Program with the purpose of verbal behaviour coding and sequential interaction analysis. We will show the graphical options that are useful to interpret data qualitatively as well as quantitatively, and how this can be used to convert qualitative data into quantitative data. We will end with one of the key features of the program: sequential analysis. This analysis is based on the essential idea that a particular event (an utterance of the respondent) partly depends on preceding events (e.g., an utterance of the interviewer). In this way, SEQUENCE VIEWER enables researchers to study relations between events at different time points and between events of different types.

Survey Quality Predictor (SQP)

Daniel Oberski; ESADE, Universitat Ramon Llull, Spain

Willem E. Saris; ESADE, Universitat Ramon Llull, Spain

Irmtraud Gallhofer; ESADE, Universitat Ramon Llull, Spain

The quality of survey questions is sometimes estimated using factor analysis, MTMM models or other psychometric models such as IRT. Most substantive researchers, however, do not have such a measurement error study available.

This means that most of the time survey research is done without any information about the quality of the questions. Not taking measurement error into account in many cases leads to highly biased results (see e.g. Biemer & Trewin 1991).

Survey Quality Predictor (SQP) is the result of a meta-analysis of 87 multitrait-multimethod experiments conducted over the years by Andrews, Költringer, Saris, Gallhofer and Scherpenzeel (Saris 2007).

The meta-analysis links the characteristics of survey questions to their quality. Among the characteristics studied are number of categories, topic under study, factual or subjective question, etc. In total more than 48 characteristics have been studied.

SQP is a computer application programme which predicts the expected reliability and internal validity of a question, based on its observed characteristics. The programme is available for free from <http://www.sqp.nl>.

References

- Biemer, P. P. and Trewin, D. (1991). "A Review of Measurement Error Effects on the Analysis of Survey Data". In Lyberg, Biemer, Collins, de Leeuw, Schwarz & Trewin (ed.), *Survey Measurement and Process Quality*. New York: John Wiley & Sons, Inc.
- Saris, W.E. and I. N. Gallhofer (2007). *Design, evaluation and analysis of questionnaires for survey research*. New York: Wiley.

2.2 Questionnaire Translation and Adaptation

Session coordinated by:

Janet Harkness; Centre for Survey Research and Methodology (GESIS-ZUMA), Germany

Ana Villar; University of Nebraska, Lincoln, United States

Presentations

1. *Instructions and Adaptation in Survey Translation: an Experimental Study* – B. Kleiner (United States), J. Bouic (United States), Y. Pan (United States)
2. *Is Cognitive Interviewing Useful for Testing Spanish-language Translations of Dietary Questions?* – B. Forsyth (United States), K. Levin (United States), A. Norberg (United States), F. Thompson (United States), G. Willis (United States)
3. *Translating Survey Questionnaires into Spanish. . . but Which Spanish?* – G. Piani-Acosta (United States)
4. *PISA Questionnaires and the PISA Questionnaire Adaptation Spreadsheet* – S. Dept (Belgium)
5. *Survey Translation Comes of Age?* – J. Harkness (Germany), A. Villar (United States)
6. *Questions About Translation from the European Social Survey* – D. Behr (Germany)

Presentations in the session “Questionnaire Translation and Adaptation “

Instructions and Adaptation in Survey Translation: an Experimental Study

Brian Kleiner; Westat, United States

Jerelyn Bouic; Westat, United States

Yuling Pan; U.S. Census Bureau, United States

In calling for faithfulness to intended meaning as a guiding principle of survey translation, Harkness and Schoua-Glusberg (1998) argue further for the need to provide translators with documentation so that they can better understand the intended readings and research aims of survey items. The authors also point out that translators should be given guidelines regarding an acceptable degree of adaptation. In actuality, however, translators are rarely provided with such documentation and guidelines, and while doing so seems reasonable in theory, there is currently little empirical work that lends support to the utility of this practice. The experimental study described in this paper examined the impact of providing special instructions and supporting material to translators. Specifically, it addressed whether translators provided with explanatory text and guidelines were able to produce translations that were more faithful to the intended meaning of source survey items and more culturally appropriate and natural sounding compared to translators who received no such guidance.

To assess the impact of different types of instructions, 27 professional translators translated an English source instrument into three target languages – Mandarin Chinese, Spanish, and Canadian French (9 translators per language) – following one of three sets of instructions (translators were randomly assigned to three instructional subgroups). The first subgroup received a core set of instructions. The second and third subgroups received the core instructions as well as question by question explanations of the intended meaning of source items, and were told to translate in a way that was faithful to the intended meaning of the source items. The third subgroup received an additional instruction to translate in a natural sounding way with respect to the cultural and linguistic norms of the target population.

Once the translations were completed, 15 professional survey researchers (5 per language) who were native speakers of the target languages conducted blind evaluations, with each examining three translated versions of survey items (that followed the three sets of instructions). The evaluation involved rating each of the translated survey items on Likert scales along three dimensions – overall quality, faithfulness to intended meaning, and cultural appropriateness. Telephone interviews were subsequently conducted with the evaluators to better understand their beliefs and practices regarding survey translation and how they approached the evaluation task.

Study findings indicate that while the provision of special instructions and documentation to translators had a significant impact on their translations, the direction of the impact (positive or negative) differed across the target languages, according to ratings of the survey researchers. We attribute these differences to the beliefs of the survey researchers and their level of commitment to two conflicting general types of equivalence in survey translation: equivalence of stimulus and equivalence of effect. Our research suggests that the issue of providing translators with documentation and guidelines is more complex than assumed, and that researchers should consider carefully in advance whether this is necessary, given the nature of the survey as well as their own beliefs about adaptation and the primacy of equivalence of stimulus or equivalence of effect.

Is Cognitive Interviewing Useful for Testing Spanish-language Translations of Dietary Questions?

Barbara Forsyth; Westat, United States

Kerry Levin; Westat, United States

Alicia Norberg; Westat, United States

Frances Thompson; National Institutes of Health (NIH), United States

Gordon Willis; National Institutes of Health (NIH), United States

Cognitive interview pretesting is generally acknowledged as an important step in questionnaire development. However, in the context of cross-cultural research, there is disagreement about the effectiveness of cognitive interviewing as a pretesting tool. For example, translation researchers have reported that respondents with different backgrounds may deal with think-aloud probes differently (e.g., Pasick, 2001; Goerman, 2006a). Also, cultural differences in direct and indirect communication styles and cultural differences in openness expressing opinions may affect the utility of cognitive interviews (e.g., Pan, 2003).

This paper reports results from research exploring whether cognitive interviews can be used to improve a Spanish-language translation of English dietary questions from the 2005 National Health Interview Survey (NHIS) conducted in the U.S. by the National Center for Health Statistics (NCHS). We will focus on identifying and illustrating the kinds of issues that cognitive interviews uncovered and the kinds of revisions identified based on cognitive interview results.

We developed the Spanish-language translation using a translation and review process described in Forsyth et al. (in press) and inspired by the TRAPD model described by Harkness et al. (2003) and the PTPRD model used by the U.S. Census Bureau. We conducted 36 cognitive interviews in several locations within the U.S. to make sure we included respondents from a variety of Spanish and Latino backgrounds. We included Spanish and English-speaking respondents to see whether similar issues emerged for both languages.

Cognitive interviews identified important shortcomings in the original questionnaires. We found translation wording problems and general problems with comprehension and recall that seemed to transcend language. We found relatively few culture-specific problems. Further analysis revealed sub-categories within each main problem area. This paper will review the various types of problems found through cognitive interviewing and illustrate them with specific examples.

The evident success using cognitive interviewing to identify problems with translated survey items is important because other researchers have reported mixed results. At least two factors probably contributed to successes reported here. First, the cognitive interview protocols followed preliminary guidelines recommended by Goerman (2006) to help volunteer respondents understand the structure and purposes of cognitive interviews. Second, Spanish-speaking cultures may be relatively close to the U.S. along the continuum of directness, a factor Pan identified as likely to be important in pretest method effectiveness. NCI is currently conducting similar research testing Asian-language translations of diet questions. Future comparisons between Spanish-language and Asian-language results using cognitive interviews may shed additional light on how these two factors affect cognitive interview results.

Translating Survey Questionnaires into Spanish... but Which Spanish?

Giorgina Piani-Acosta; University of Michigan, United States

Spanish, like other romance languages, derives from Vulgar Latin, the cluster of dialects spoken by legionaries, traders and farmers in the Roman Empire. In America, the history of Spanish starts with the Spanish colonization at the end of the fifteenth century. During this period, the southern part of the American continent was a conglomerate of hundreds of different indigenous languages and dialects.

The conquest and colonization of Central and South America (excluding the territory of Brazil) was accompanied by a slow but firm imposition of the Spanish language. However, this process was a two-way flow of cultural and linguistic influence between the colonizers and the colonized that allowed the incorporation of aspects belonging to the pre-Columbian cultures into what would later become American Spanish. African slaves and European immigrants in the late 19th and early 20th century also played a critical role in the formation of the American Spanish.

The magnitude of the territory and the history of the different geographical areas in Latin America paved the way for the development of different variants of the Spanish brought from Spain. These variants are called “regionalisms” and involve not only intonation and pronunciation differences, but also vocabulary differences (words and expressions). Nowadays, at least seven Spanish Regionalisms

can be identified: (1) U.S. Spanish (2) Caribbean (3) Mexican (4) Andean (5) Chilean and (7) River Plate.

When planning to translate a source questionnaire into Spanish, special attention should be paid to these differences. As the failure to produce easy to understand questions that speak the same language to the respondent might seriously harm the validity of the measures.

Specifically, bizzard formulated questions and/or weird response categories could affect respondents reporting through the following mechanisms:

1. misinterpretation of the question
2. failure to follow instructions accurate
3. interviewer‘difficulty to build rapport with the respondent
4. need of interviewer‘clarifications of the question

But, to what extent the use of a “non local” Spanish affects the validity of the instrument? If it does, in what direction is it affected? What specific countries are more sensitive to regionalism differences? What special cares should we take into account?

To try to answer these questions, a set of randomized experiments were fielded in Uruguay with the 2006 International Social Survey Program (ISSP) questionnaire on Role of Government IV. Two sets of questions were selected from four different country Spanish versions (Spain, Dominican Republic, Mexico and Uruguay) translated from the same ISSP English source.

PISA Questionnaires and the PISA Questionnaire Adaptation Spreadsheet

Steve Dept; cApStAn Linguistic Quality Control, Belgium

The paper will describe the methodology used to discuss, document, approve and monitor national adaptations of contextual Questionnaires within the framework of the OECD/PISA 2006 Survey cycle. A number of concrete examples of adaptation difficulties and ways to resolve them will be presented.

Participating countries listed their national adaptations (all desirable or required deviations from the international source version) in a Questionnaire Adaptation Spreadsheet (QAS), together with a back-translation into English and a justification.

A team of experts considered the adaptations and agreed, or did not agree (giving reasons), or requested further information regarding each adaptation.

This was an iterative process of communication between the National Centre of the participating country and the International Study Centre. Following the resolution of all issues related to the adaptation of the Questionnaires, an agreed QAS was returned to the National Centre.

The National Centre submitted to the international verification team (cApStAn) the “agreed” QAS and the national versions of their Questionnaires, aligned to the agreed QAS.

The verifiers’ mission was to:

1. verify linguistic aspects of the national version of the Questionnaires;
2. verify the correct implementation of approved adaptations listed in the QAS; and
3. identify possible adaptations or deviations not listed in the QAS.

The verifiers systematically checked whether the national version of the adaptation was equivalent to its English back-translation, which the consortium had “agreed”. Verifiers then checked whether those adaptations were correctly implemented in the Questionnaires themselves. If a verifier identified a mismatch between what was agreed in the QAS and what was stated in the Questionnaire, s/he aligned the Questionnaire so that it would match the QAS.

The verifiers also checked the Questionnaires for linguistic correctness as well as undocumented adaptations [deviations versus the source text that were not listed in the QAS]. They entered corrections

in track changes in the MS Word Questionnaires and entered comments in the verifier column of the QAS.

Instructions to the verifiers were straightforward, and the procedure proved to be effective. Verifiers refrained from discussing agreed adaptations unless the back translation into English of the agreed adaptation inadequately conveyed its meaning, in which case the Consortium might unknowingly have approved an inappropriate adaptation.

The verified Questionnaires were returned together with an annotated QAS, which was also sent to the International Study Centre. The National Centre edited the questionnaires according to the comments received from the verifier. Simultaneously, staff from the International Study Centre highlighted verifier interventions that required follow-up. Such issues were earmarked for verification at Final Optical Check stage.

After implementation of corrections suggested by the verifier, pdf copies of the finalised national versions of the questionnaire instruments were sent to the verification team for a “Final Optical Check” (i.e. verification of the layout, instructions to the student, etc). This included checking whether controversial issues were solved in a satisfactory way.

Survey Translation Comes of Age?

Janet Harkness; Centre for Survey Research and Methodology (GESIS-ZUMA), Germany

Ana Villar; University of Nebraska, Lincoln, United States

The introductory paper to the translation session presents what we see as the important recent innovations and trends in survey research against the ongoing challenges faced. In doing so, we attempt to provide a backdrop for each contributing paper. We begin with the relation between source question design and translation, and to the role of adaptations in producing other language version of source questionnaires. We turn then to innovations in translation production and translation review, in particular considering team approaches to translation and translation review viewed against other popular approaches to translation. In this context, we also consider the place of back translation assessment in survey research.

The large projects becoming increasingly common in comparative survey research often have several countries fielding in the same language(s). Examples are readily found both in surveys within Europe and those further afield. We explore the challenges and opportunities offered though countries sharing languages and present different models for dealing with this fact.

New technological developments and the increased emphasis placed on documentation in modern translation procedures have highlighted the need to integrate survey translation into survey process management systems. The paper reviews these developments and the emergence of other support devices and services for project translation efforts (e.g., the support desk in the ESS, the translation work group in the ISSP). We conclude with comments on a number of prominent guidelines on translation that are available in different projects or disciplines and a review of recent research initiatives as these relate to survey translation production and quality.

Questions About Translation from the European Social Survey

Dorothee Behr; Centre for Survey Research and Methodology (GESIS-ZUMA), Germany

High-quality questionnaire translations are a prerequisite for data comparability in multilingual survey research. However, translating questionnaires is not as straightforward and simple as it may seem. Questionnaires are both special kinds of texts and tools of measurement and this turns questionnaire translation into a complex undertaking.

The European Social Survey (ESS) is quite unusual in the approach it takes to translation. Not only are participating countries provided with guidelines on how to conduct and review their translations, they also have the opportunity to approach a “help desk” with queries they may have during the course of translation. The people behind this virtual help desk are basically researchers working at ZUMA,

Mannheim, Germany and researchers attached to the ESS central co-ordinating office in London. They in turn contact other experts if and as the need arises.

Against the background of the general support system for the translation process in the ESS the paper presents and discusses the questions asked by countries translating during Round 3 of the ESS. It presents what we might learn from these "FAQ" with regard to translations, translators, the information countries may lack and, ultimately, also source questionnaires. The aim is to use insight gained from analysis of why questions are asked and where and how misunderstandings occur to work towards a better and pro-active training and briefing of translation teams.

2.3 Questionnaire Development and Testing

Session coordinated by:

Manuela Murgia; ISTAT, Italy

Presentations

1. *Testing the Stability of an Acquiescence Style Factor Behind Two Interrelated Substantive Variables in a Panel Design* – J. Billiet (Belgium), E. Davidov (Germany)
2. *How to Analyze Data Collected Using the Semantic Differential Technique* – O. Mitina (Russia)
3. *Factor Analysis with Ordinal Data: an Investigation on the Impact of the Number of Response Alternatives of the Manifest Variables* – K. Kraus (Sweden)
4. *Standardized Methods for Developing Questionnaires: an Overview of the Standardization Process at Statistics Sweden* – A. Persson (Sweden)
5. *Effects of Different Layouts of an Electronic Calendar Recall Aid* – T. Glasner (Netherlands), W. van der Vaart (Netherlands)
6. *Effects of Respondent Focus in Questionnaire and Data Collection in a Public Health Survey* – M. Wenemark (Sweden), M. Kristenson (Sweden), H. Noorlind-Brage (Sweden)
7. *Deciding Whether the New is Better Than the Old: the Role of Multiple Evaluation Methods in Assessing the Performance of a National Travel Diary* – D. Collins (United Kingdom), A. McGee (United Kingdom)
8. *New Experiments on the Optimal Structure of Complex Survey Questions* – P. Beatty (United States), F. Fowler (United States), C. Cosenza (United States)

Presentations in the session "Questionnaire Development and Testing"

Testing the Stability of an Acquiescence Style Factor Behind Two Interrelated Substantive Variables in a Panel Design

Jaak Billiet; University of Leuven, Belgium

Eldad Davidov; University of Köln, Germany

This article addresses the question to what extent one type of response style called acquiescence (or agreeing response bias) is stable over time. Several options in order to style effects are discussed. A Structural Equation Modelling (SEM) approach is then applied to measure the stability of one acquiescence factor behind two concepts among the same respondents over a four-year period of time. The data used are representative population surveys in 1995 and 1999 from the Belgian Election Study (N=1,112) in which balanced sets of items are used for measuring two interrelated constructs, perceived ethnic prejudice and political efficacy. This study provides empirical support that acquiescence is stable and consistent over a four-year period of time. It is shown that the measurement of response styles is useful in cross-cultural research.

How to Analyze Data Collected Using the Semantic Differential Technique

Olga Mitina; Moscow State University, Russia

Semantic differential itself and its different variants are useful and relevant methods of data collection in research of implicit stereotypes including cross-cultural studies. Developing three mode analysis data is very important in these cases. Researchers using the well known PARAFAC and three mode factor analysis offered by Tucker meet with difficulties with the interpretations of results. That is why alternative methods of three-mode data analysis will be suggested; we can expect more successful analysis and interpretation (especially if we produce several different data analysis and compare them with each other).

In this paper we suggest one way of data collection using semantic differential technique, when the subjects evaluate different concepts using a set of scales. For each subject we calculate a vector of numbers. Each of these number is a correlation between a pair of concepts (vectors of its evaluations by the set of scales). Then we get groups of correlation among concepts using usual method of factor analysis. Some of correlations among different pairs of the concept are in the same group if they have high factor loadings on the same factor among extracted ones. To interpret results we use graphs for each extracted factor. The vertexes of graph are the concepts. To vertexes are connected by edge if the pair of the corresponding concepts have a high loading on the corresponding factor. As result we can describe subjects which got high factor scores on this or that factor as persons who have image of relations among scaled concepts characterizing by high connectivity the concepts connected by edges in the graph. This visual representation can be useful when we study self conception and the concepts are different parts of Myself: "Real self", "Ideal self", "reflected Self" (as I am seen by others), "Due self" (How I should be), "Neglected Self". Another example which also will be demonstrated is subjective image of the world. The subjects scale images of different countries and according of different configurations of the countries we can describe different patterns of thinking about countries coalitions and global similarities.

Factor Analysis with Ordinal Data: an Investigation on the Impact of the Number of Response Alternatives of the Manifest Variables

Katrin Kraus; University of Uppsala, Sweden

During the last years, researches in social sciences have become more conscious regarding the issue of ordinal data in quantitative analysis. A range of methods have been developed to take the properties

of ordinal data into account. However, some apparently simple issues regarding the construction of questionnaires remain unsolved until today.

The present study investigates the influence of the number of categories of ordinal variables on parameter estimates and inference in factor analysis with robust maximum likelihood that uses polychoric correlations. A simulation study is conducted to examine eventual effects on the bias of parameter estimates, their standard errors and goodness-of-fit statistics. In addition to the number of response alternatives (2, 3, 4, 5, 7, 10, and 15), the distribution of the responses on the categories is varied systematically to investigate the impact of skewed or heavy tailed data. Different compositions of the ordinal indicators are considered for two sample sizes ($N=80, 500$). For all cases, special attention is paid to the issue of dichotomization of ordinal manifest variables.

The results indicate that the number of categories is of little importance for the inference in factor analysis for the larger sample size and for situations with symmetrical items. For the smaller sample size, problems are obtained for the smaller number of categories (2 and 3) especially for highly skewed manifest variables. Occurring problems concern mostly the rate of convergence and non admissible parameter estimates. To summarize, the present study results in important implications for the construction of questionnaires that are supposed to be analyzed by factor analysis. Besides the rather obvious recommendation of large sample sizes and not too skewed items, the results show that the negative effect of small sample sizes can be compensated by a larger number of categories.

Standardized Methods for Developing Questionnaires: an Overview of the Standardization Process at Statistics Sweden

Andreas Persson; Statistics Sweden, Sweden

The questionnaire is one of the most common instruments to gather information. However, flaws in the design can have severe consequences on the data quality. As such, it is crucial to improve and test the questionnaire prior to data collection in order to minimize measurement error. To guide us there are a number of different qualitative and quantitative methods at hand, including expert reviews, cognitive interviewing, focus groups, in-depth interviews, usability tests and formal experiments. The question is, however, when and how to apply these methods?

A standardization project at Statistics Sweden is developing standard procedures and guidelines for the applications of these methods. We take a closer look at a number of specific cases and classify them in factors that are relevant when choosing method (such as mode of data collection, single or multi-round survey etc). Then, we determine which method, or combination of methods, that is the most appropriate in the given case. In this way, we aim to establish type cases in questionnaire development and guidelines for practice. In addition, this work gives insights into which factors that tend to influence the choice of methods.

Another aim of the project is to establish detailed standard procedures for each method. Such standards will entail all stages in the procedure, such as the selection of test sample, cognitive interviewing guidelines, probing techniques, analyzing routines, documentation etc.

This presentation present the work done in the project.

Effects of Different Layouts of an Electronic Calendar Recall Aid

Tina Glasner; Vrije Universiteit Amsterdam, Netherlands

Wander van der Vaart; Vrije Universiteit Amsterdam, Netherlands

In this study, we used a cognitive interviewing technique to compare the usability of different versions of a retrospective questionnaire including a calendar recall aid, to be used in a large-scale online survey. The aim of our study was twofold: Firstly, we were asked for suggestions concerning the optimal design of the calendar recall aid. Calendar instruments are currently used in many retrospective surveys, and have been found effective in reducing recall error. Even though several electronic versions have

been developed based on the original paper-and-pencil method, those instruments are interviewer-administered, and too complicated to be included into a web-based survey. We are therefore facing the challenge of simplifying the recall aid and adjusting it to the mode of data collection, without compromising its beneficial effects on data quality. Secondly, we were interested in the memory processes involved in answering retrospective questions with or without a calendar recall aid. The effectiveness of calendar recall aids has been theoretically linked to hierarchical models of autobiographical memory. The present study sets out to find empirical evidence on this connection.

In order to fulfil both research objectives, we conducted 37 cognitive interviews in three experimental conditions. Respondents who were recruited from an existing online panel were asked to fill out either a traditional-style retrospective questionnaire without a recall aid, the same questionnaire combined with a (paper-and-pencil) calendar, or a flexible calendar questionnaire. Respondents were encouraged to think aloud, while completing the test questionnaires. After the interview, test subjects were asked for feedback concerning the layout of the questionnaire and the difficulty of the retrospective questions asked during the interview. We made audio and video recordings of the interviews, which were transcribed and then coded using Sequence Viewer 4. With regard to recommendations for optimal design our focus was on the order of domains, forward or backward recall strategies, and the types of landmarks (pre-specified or idiosyncratic) used by respondents. As to our second research goal, we coded the memory processes respondents mentioned in their verbal reports. Since all our test subjects were members of an online panel, we will be able to compare some of their responses to their own concurrent reports from earlier waves of the panel study.

Effects of Respondent Focus in Questionnaire and Data Collection in a Public Health Survey

Marika Wenemark ; Linköping University , Sweden

Margareta Kristenson; Linköping University , Sweden

Helle Noorlind-Brage; Centre for Public Health Sciences, Östergötland County Council, Sweden

Public health questionnaires have, in Sweden, a long tradition as important means of monitoring populations' health. Nonresponse is a growing problem in these surveys just as it is in many others. Improving response rates by persuasion of reluctant respondents can result in reduced data quality and might deteriorate the respondents' willingness to participate in future studies. In a survey regarding mental health 2002, methods focusing on respect for the respondent were used as an alternative to intensified follow-up.

Based on the positive experiences from that study, an experimental design was included in the 2006 public health survey in Östergötland County with a total sample of 16 940 persons. Three different combinations of respondent focus in questionnaire and data collection method were compared to a typical Swedish public health questionnaire and a standard practise of data collection conducted by Statistics Sweden. The intention of the respondent focus was that respondents should feel that they could give valuable information in a truthful way without feeling persuaded, annoyed or manipulated. Respondent focused questionnaire means in this study for example fewer questions and more opportunities to give open-ended answers. Respondent focused data collection means in this study for example pre-notification letter, no persuasion in remainder letter, only one remainder and partial anonymity.

Results regarding response rate, data quality (for example item response) and respondent satisfaction (for example respondents judgement of questionnaires) will be discussed for the different versions of the experiment. The results suggest that it is possible to perform health surveys with improved response rates and higher respondent satisfaction.

Deciding Whether the New is Better Than the Old: the Role of Multiple Evaluation Methods in Assessing the Performance of a National Travel Diary

Debbie Collins; National Centre for Social Research - NatCen, United Kingdom

Alice McGee; National Centre for Social Research - NatCen, United Kingdom

Cognitive interviewing methods are a useful tool in evaluating the measurement performance of survey instruments. They provide insight into the causes and nature of problems respondents encounter when attempting to answer questions or complete a questionnaire, and in so doing may suggest possible remedies. However findings from cognitive interviews in and of themselves are often not sufficient to convince funders that they should make changes to existing survey instruments. Additional evidence is often required. One reason for this is that cognitive interview data are qualitative in nature and as such they are unable to provide statistically reliable information on the size or extent of such problems or the success of alternatives in reducing them. Split ballot experiments can, if well-designed, provide such quantitative evidence yet funders (in the UK at least) are often reluctant to fund them. So how can we demonstrate that changes made to question wording and/or document design lead to improvements in data quality?

This paper is based on a study that sought to review and revise a travel diary used on the National Travel Survey, and the evaluation strategy we developed to assess whether the new diary we designed was 'better' than the existing one. The strategy was built on the premise that we should design the revised diary by applying theoretical principles supported by empirical evidence, and that these revisions should therefore lead to measurable (quantifiable) improvements in data quality.

A number of sources of evidence were reviewed to assess the performance of the existing diary (stage one), which informed its redesign. These included: looking at levels of item missing data; discussions with interviewers and office editors on the problems they encountered; and cognitive interview data. The revised diary was then subjected to further cognitive testing and field piloted (stage two). A split ballot experiment was not possible so we developed an alternative approach, which involved the construction of a code frame that identified particular types of errors made by respondents, informed by our findings at stage one. A sample of completed existing travel diaries was coded using this frame, as were the completed 'new' diaries that were field piloted. Results were compared and triangulated with evidence from the cognitive interviews on the performance of the new diary. In this paper we: (a) discuss the development of this strategy, specifically the code frame used to compare the existing and new diaries (b) present findings from the evaluation of the two diaries; and (c) consider what we gained from adopting this multi-method approach.

The National Travel Survey is a large, cross sectional survey designed to provide regular, up-to-date data on personal travel and monitor changes in travel behaviour over time within Great Britain. Information is collected throughout the year from a randomly selected sample of households via a face-to-face interview. In addition each household member (adult and child) is asked to complete a seven day Travel Record (diary). The survey has been continuously running since 1988.

New Experiments on the Optimal Structure of Complex Survey Questions

Paul C. Beatty; National Center for Health Statistics, United States

Floyd J. Fowler; Center for Survey Research (UMass Boston), United States

Carol Cosenza; Center for Survey Research (UMass Boston), United States

Survey researchers often need their questions to convey very specific information to respondents—for example, questions may include complex definitions, instructions to include or exclude various considerations while answering, and a particular set of closed-ended responses. Although questionnaire design principles provide some advice on constructing complex questions, little empirical evidence demonstrates the superiority of certain decisions over others. For example, in some questions, important respondent instructions “dangle” after the core question has been asked; one alternative is to provide such definitions before asking the core question.

The authors have conducted several rounds RDD telephone surveys with split-ballot experiments to explore such issues. This paper reports on the latest round of 425 interviews conducted via an RDD telephone survey, in which respondents received alternative versions of various survey questions. For example, in some experiments, alternative questions used the same words but were structured differently. Other experiments compared the use of examples vs. definitions to explain complex concepts, compared the use of one vs. two questions to measure the same phenomenon, and compared questions before and

after cognitive interviews had been used to clarify key concepts. With permission, interviews were tape recorded and behavior-coded, making it possible to compare various interviewer and respondent difficulties across question versions, in addition to comparing differences in response distributions.

Taken in conjunction with findings from previous rounds of experiments, the results begin to suggest some general design principles for complex questions. For example, the disadvantages of “dangling qualifiers” are becoming clear, as are the advantages of using multiple questions to disentangle certain complex concepts. The paper will report results of these and other experimental comparisons, with an eye toward providing more systematic questionnaire design guidance.

2.4 Communicating Survey Quality to Users

Session coordinated by:

George Petrakos; Agilis SA, Greece

Presentations

1. *Quality of the Statistical Information: Important Precondition for the Reliability of Global Economic Analysis* – V. Pavlova (Bulgaria)
2. *The Metadata of Frames, Populations and Coverage As a Quality Assessment Tool in Statistical Surveys* – G. Petrakos (Greece), S. Chrysanthopoulou (Greece)
3. *Assessing the Value of a Survey* – T. Chen (United Kingdom), R. Raeside (United Kingdom)

Presentations in the session “Communicating Survey Quality to Users”

Quality of the Statistical Information: Important Precondition for the Reliability of Global Economic Analysis

Veselka Pavlova; University of National and World Economy, Bulgaria

In the current process of Euro integration and globalization, a main task for every statistical institution is to provide sufficient information with good quality as a reliable base for national and international analysis. Concerning users' satisfaction with the quality of the data produced and provided by the National Statistical Institute (NSI) of Bulgaria a specialized study was carried out in August and September 2005 . It was based on a summary of the subjective opinions, assessments and recommendations of 1502 users. They were distributed in several basic consumers' segments – business enterprises, central public administration, district administration, municipal administration, non-state organizations and educational institutions. The components refer to the quality recommended by Eurostat are: accessibility; relevance; data juxtaposition; data reliability; terms for information provision; accompanying analyses and assessments (metadata) and clarity of the indicators used. General statistical programs like SPSS or STATA, that offer very good tools for data analysis, starting from simple descriptive statistics were used.

On the basis of the survey's results and their analysis following main conclusions can be drawn:

1. The majority (over 75%) of the users of statistical information produced by the NSI, is concentrated in the big residential areas and the country's capital;
2. Over 1/3 of the interviewed persons use statistical information, where 11% use it constantly and 14% use it often;
3. Regular statistical information users are mainly (about 62%) structures related to the state administration, health care and social work and education. Those who use very rarely or do not use at all statistical information are primarily from business enterprises (42 %);
4. There are two main reasons for an organization's not using NSI's data: first the interviewed think that such information is not necessary for their organization (48%) and second is that it doesn't provide the information it needs (36%) itself;
5. Users' satisfaction with the provided statistical information is evaluated by way of two prisms: users' satisfaction with the information servicing provided by the NSI and assessment of separate statements connected with the statistical data's quality. The analysis demonstrates that more than 82% of the interviewed persons are completely or in most cases satisfied with the quality and only 1,4% are entirely dissatisfied;
6. The assessed average grades for the quality components of the statistical data vary from 4,40 to 5,28;
7. The highest assessment is given for the “accessibility” component (47%), followed by the “data reliability” (41%) and “clarity of the indicators used” (38%). The lowest evaluation receives the component “accompanying analyses and assessments” (15%);
8. The main directions in which the users would like to make changes in the NSI's information are: a) the terms for preparation and provision of the statistical information, b) the information's punctuality and reliability, c) the coverage of the indicators used and the d) juxtaposition of the statistical data.

The Metadata of Frames, Populations and Coverage As a Quality Assessment Tool in Statistical Surveys

George Petrakos; Agilis SA, Greece

Stavroula Chrysanthopoulou; Agilis SA, Greece

The target population and the sampling frame, also called register, are two important building blocks of almost every statistical survey. The clear definition and determination of the population under study and the register in hand as well as a qualitative and quantitative study of the coverage error are key issues for the assessment and improvement of the quality of any statistical survey. In this paper an effort has been made to summarize and codify these key issues as metadata elements, which then can be part of the overall survey metadata vector. This metadata codification can be used along with appropriate statistical methods as the basis for assessment and comparison of related quality issues among surveys executed in different time and place.

Key Words: Target Population, Sampling Frame, Metadata elements, Coverage Error, Quality of Statistical Survey.

Assessing the Value of a Survey

Tao Chen; Napier University Business School, United Kingdom

Robert Raeside; Napier University Business School, United Kingdom

More and more surveys are becoming available to social researchers, who have to make judgments as to the efficiency, efficacy and effectiveness of the particular survey to assist in their research. In this paper we report on preliminary work to design a tool to give a holistic assessment to the researcher as to the value of the particular survey. This tool is essentially an excel based audit following through all stages of the survey from inception, to design, construction, execution, analysis and dissemination. At each stage a set of criteria are applied to make an assessment of quality. Scores attained against these assessments are weighted by the importance of the criteria and summed to give an overall assessment of the stage. Although essentially qualitative in nature the aim is to continue development and form this into a more quantitative based assessment. The real advantage of this is to construct a means of survey assessment which can be used in a diagnostic manner to improve survey quality. How to calibrate the instrument for survey size and scope will be given consideration.

2.5 Interviewer and Respondent Behaviour in Survey Interviews

Session coordinated by:

Wil Dijkstra; Vrije Universiteit Amsterdam, Netherlands

Yfke P. Ongena; Vrije Universiteit Amsterdam, Netherlands

Presentations

1. *The Effect of the Question Topic on Interviewer Behaviour; an Interaction Analysis of Control Activities of Interviewers* – J. van der Zouwen (Netherlands), J. Smit (Netherlands), S. Draisma (Netherlands)
2. *Evaluation of Coding Schemes with Direct Informants and Proxies* – I. Benítez Baena (Spain), J. Padilla Garcia (Spain)
3. *Cross-sectional and Longitudinal Interviewer and Respondent Effects in a CATI Panel Survey* – O. Lipps (Switzerland)
4. *Practice of Respondent Selection and Sampling Quality in Surveys* – N. Menold (Germany)
5. *Behaviour Coding of Face-to-face Interviews with Children: Question Difficulty and the Impact of Cognitive Resources on Response Quality* – M. Fuchs (Germany)
6. *Applying Behaviour Coding to Monitor Interviewer Performance in Event History Calendar Interviews in a Cross-national Survey* – W. Dijkstra (Netherlands), Y. Ongena (Netherlands)

Presentations in the session "Interviewer and Respondent Behaviour in Survey Interviews"

The Effect of the Question Topic on Interviewer Behaviour; an Interaction Analysis of Control Activities of Interviewers

Johannes van der Zouwen; Vrije Universiteit Amsterdam, Netherlands

Johannes H. Smit; Vrije Universiteit Amsterdam, Netherlands

Stasja Draisma; Vrije Universiteit Amsterdam, Netherlands

In a standardized personal interview, elderly (65+) Dutch respondents (N=233), were asked detailed retrospective questions about six physical activities like walking, cycling and their performance of household tasks. Surprisingly, the proportion of inadequate answers was small, suggesting that the interviewers – four professional research nurses – have done their very best to eventually obtain adequate answers. They used three different types of control of the interview process. Firstly, an optimal execution of the prescribed 'open loop control', that is, precisely following the text of the questionnaire. Secondly, via the 'feedback loop control' of repair: if nevertheless inadequate answers are given, further probing is performed until eventually an adequate answer is obtained. Thirdly, by decomposing a general question into its components, making the question easier to answer (partial questioning).

Interaction analysis of transcripts of the interviews showed that the type of control exercised by the interviewers, strongly differed by topic of the question. It appeared that question topics requiring more complex cognitive activities of the respondents, not only lead to more (need for) repair, but also to larger differences between the interviewers concerning their readiness to use partial questioning as an alternative for the open loop control as designed by the researcher.

Keywords: question topic; retrospective questions; survey interviews; interviewer related measurement error; interviewer behaviour; behaviour coding, interaction analysis

Evaluation of Coding Schemes with Direct Informants and Proxies

Isabel Benítez Baena; University of Granada, Spain

Jose Luis Padilla Garcia; University of Granada, Spain

"Behaviour coding" has proved to be a useful procedure to detect potential sources of measurement errors in questionnaire design. The quality of the coding scheme can determine the validity of the information drawn when analyzing the behaviour of the interviewer, and the respondent, as well as the interaction between both.

Studies are scarce on the adequacy of the coding schemes when the respondents act as either indirect informant or "proxy". Household-based surveys frequently request "proxy" informants. Proxies' responses are used for complete enumeration of the household and for establishing the eligibility of certain household members for particular questions. The proxy's answer can determine the quality of the information obtained from the survey.

The goal of this work is to examine the utility of the coding scheme proposed by Dijkstra and Ongena (2006) in order to compare the behaviour of direct informants with that of proxies. We present here part of the results of the cognitive pretest of a questionnaire of disabilities. 25 direct informants and 25 proxies responded to 11 "target" questions which are the same for both types of informants. Finally, the convergence between the information yielded by behaviour coding and the cognitive interview procedure is discussed.

Cross-sectional and Longitudinal Interviewer and Respondent Effects in a CATI Panel Survey

Oliver Lipps; Swiss Household Panel, Switzerland

Especially in panel surveys, respondent attrition, respondent learning, and interviewer experience effects play a crucial role with respect to data quality. We examine three interview survey quality indicators in the same survey in a cross sectional as well as in a longitudinal way. In the cross sectional analysis we compare data quality in the mature original sample with that in a refreshment sample, surveyed in the same wave. Because in the same wave an interviewer survey was conducted, collecting attitudes on their socio demography, survey attitudes and burden measures, we are able to consider interviewer fixed effects as well. The longitudinal analysis gives more insight in the respondent learning effects with respect to the quality indicators considered by considering the very same respondents across waves.

The Swiss Household Panel, a CATI survey representative of the Swiss residential population, forms an ideal modelling database: the interviewer – respondent assignment is random, both within and across waves. This design avoids possible confusion with other effects stemming from a non-random assignment of interviewers, e.g. area effects or effects from assigning the best interviewers to the hard cases. In order to separate interviewer, respondent and wave effects, we build cross-classified multilevel models.

Practice of Respondent Selection and Sampling Quality in Surveys

Natalja Menold; Centre for Survey Research and Methodology (GESIS-ZUMA), Germany

Careful sampling design in surveys is performed to obtain high quality data. The European Social Survey (ESS, Round 1) established strict random procedures, standards, and documentation. The control of real fieldwork is still problematic in particular countries. In the presentation it will be shown that interviewers influence actual participants' selection. The data quality of ESS in different countries was tested by means of internal criteria by paired comparison. The main point of this analysis is to derive a subpopulation in which the distribution of a parameter is known. This parameter is gender, if only gender homogeneous couples living in a household are separated. A significant deviance in the sample of the expected gender distribution (50% males and 50% females) denotes accessibility effects. The distribution of the gender parameter was analysed in derived subpopulations: couples with children and older married persons living in a household. Significant differences in interviewing of male and female persons in these groups were found in many countries, depending on the factors job, children and role difference. The results indicate a systematic influence of the behaviour of the interviewers and survey participants. Countries with different sampling design are compared to show design effects. How such biases influence the cross-cultural comparability of particular respondent attitudes was tested on different constructs (e.g. ethnic prejudice, political participation). Finally, the possibilities of controlling fieldwork to improve sample quality, for example by following documentation standards established in ESS, are discussed.

Behaviour Coding of Face-to-face Interviews with Children: Question Difficulty and the Impact of Cognitive Resources on Response Quality

Marek Fuchs; University of Kassel, Germany

Increasingly, children of all ages are becoming respondents in survey interviews. While juveniles are considered to be reliable respondents for many topics and survey settings it is unclear to what extent younger children provide reliable information in a face-to-face interview. So far, findings are mostly based on field experimental studies assessing the quality of the responses and the underlying cognitive processes using measurement error indicators derived from self administered surveys.

In this paper we will report results from a study using video captures of 150 face-to-face interviews with children aged 9 through 11. Interviewers were trained to apply either standardized or flexible interviewing techniques. All interviews have been coded using behaviour codes on a question by question level which provides behaviour-related indicators regarding the question answer process. In addition, standard tests of cognitive resources have been conducted.

Based on a multi-level analysis cognitive resources of the young respondents as well as interviewer style, question difficulty and question style will be discussed as factors determining the quality of the response obtained from children in standardized interviews.

Applying Behaviour Coding to Monitor Interviewer Performance in Event History Calendar Interviews in a Cross-national Survey

Wil Dijkstra; Vrije Universiteit Amsterdam, Netherlands

Yfke P. Ongena; Vrije Universiteit Amsterdam, Netherlands

The most objective and reliable way to monitor interviewer performance in face-to-face interviews, is to audio record the interviews and systematically code the interviewer's behaviour with respect to the aspects that are most important to the study. Some authors however (e.g. Haunberger, 2006) find this method hardly feasible, because of ethical and financial reasons.

The present study shows that behaviour coding is very well feasible, even in a very complex study. We applied behaviour coding in a large scale epidemiological study among 26,000 patients from hospitals and clinics from a number of different European countries (e.g., Germany, France, Italy and Greece). The study makes use of an Event History Calendar (EHC) to gather data about many aspects of the respondent's life (residences, marriages, birth of children, education, jobs, smoking and drinking history, diet, medical history). The interviews take on average about two hours, and are administered in the respondent's native language. All interviews are audio recorded (conditional on the respondent's permission) and more than 10 percent of these recordings are subjected to behaviour coding. Behaviour coding is focused on the adequate reading of introductions and questions, and the appropriate application of EHC principles, like cross-checking and probing for changes. Interviewers are provided with feedback, based on the results of the behaviour coding, within a few days after they administered the interview. If their performance remains systematically below an acceptable level, they are retrained or withdrawn from the study. The paper discusses how we implemented the behaviour coding procedures and the logistics related to evaluating the interviewer performance and providing feedback. Results will be discussed with respect to the data available at the time of the conference (collecting data started about a year ago and will continue until the end of 2008).

2.6 Augmenting Survey Data by Bio-indicators, Geo-data and Record-linkage

Session coordinated by:

Rainer Schnell; University of Konstanz, Germany

Presentations

1. *Improving Record-linkage-software for Survey-data* – R. Schnell (Germany), T. Bachteler (Germany), J. Reiher (Germany)

Presentations in the session "Augmenting Survey Data by Bio-indicators, Geo-data and Record-linkage"

Improving Record-linkage-software for Survey-data

Rainer Schnell; University of Konstanz, Germany

Tobias Bachteler; University of Konstanz, Germany

Jörg Reiher; University of Konstanz, Germany

Increasingly survey data is being enhanced by individual administrative data. Such linkages may be used to improve data quality of the surveys. An example are work histories, because respondents tend to underreport short spells of unemployment. In order to perform such linkages two problems have to be solved: Data protection objections and error prone common identifiers. The first problem can be solved by explaining the purpose of the linkage to the respondents, the second problem can be solved by using special software. The presentation will give details on current consenting procedures and consenting rates in German population surveys. Furthermore, technical improvements of record linkage software will be demonstrated. These improvements are based on enhancing the identifiers (names, addresses) with other information. These additional information may be used to reduce the number of comparisons among similar identifiers of respondents ("blocking"). Examples for such information enhanced blocking procedures are a Bayes-Classifer for nationality and a list of names of birth-places to reconstruct migration histories. Finally, since most respondents move within short distances, we compiled geo-coordinates for about 15.000 post codes in Germany. So we can use geographical distance between two cases either as a blocking variable or as part of the distance function. Finally, the application of these procedures will be demonstrated in recent research projects.

2.7 Improving the Validity of Psychological Questionnaires in the Context of Survey-based Research

Session coordinated by:

Jose Luis Padilla Garcia; University of Granada, Spain

Juana Gómez Benito; University of Barcelona, Spain

María-Dolores Hidalgo ; University of Murcia, Spain

Presentations

1. *How to Estimate and Control for Order Effects Using Generalized IRT Models: a Feasibility Study* – H. Matschinger (Germany), R. Alexandrowicz (Austria)
2. *How to Estimate and Control for Order Effects Using Generalized IRT Models: Prerequisites and Limits* – R. Alexandrowicz (Austria), H. Matschinger (Germany)
3. *Satisfaction Surveys of External Clients As a Tool to Identify and to Implant Areas of Improvement in the Public Health System* – P. Tejero Cabello (Spain), J. Gómez Benito (Spain), J. Padilla Garcia (Spain), S. Juncosa (Spain), M. Martín (Spain), S. Nofuentes (Spain)
4. *Utility of Polytomous Differential Item Functioning Techniques in Cross-cultural Survey Data* – M. Hidalgo (Spain), J. Gómez Benito (Spain)
5. *Investigating Measurement Invariance of the Infant-toddler and Emotional Assessment (ITSEA) Scale Across Demographic Groups in the Netherlands* – F. Galindo Garre (Netherlands), M. Crone (Netherlands), R. Stoel (Netherlands)
6. *Direction of Response Categories, Specificity of Items and Response Behavior* – D. Krebs (Germany), J. Hoffmeyer-Zlotnik (Germany)

Presentations in the session "Improving the Validity of Psychological Questionnaires in the Context of Survey-based Research"

How to Estimate and Control for Order Effects Using Generalized IRT Models: a Feasibility Study

Herbert Matschinger; University of Leipzig, Germany

Rainer Alexandrowicz; University of Klagenfurt, Austria

In order to measure attitudes by means of latent variable models sets of stimuli (items) quite frequently are employed invariantly over groups of individuals which may be different with respect to many characteristics. Of course, it is implicitly assumed that the effects of these potential predictors do not depend on the items employed, and therefore have not to be estimated for each item separately. On the other hand, these characteristic may cause considerable Differential Item Functioning (DIF) or item bias. One important, but almost neglected source of bias is the order of the items employed to measure the latent construct. Effects on the response for a particular item may result from learning or tiring processes and/or may due to the fact that the respondents can not be characterized by a latent, true value for the domain under consideration. This is particularly true for item-sets commonly administered in population surveys, since respondents are not familiar with the domain under investigation, but rather learn about the very topic in the course of responding to the questions.

Like other forms of item bias, the effect of the order of items can only be investigated by means of a model, which parsimoniously represents the data at hand. Only one-dimensional models will be considered here. To estimate the order effects a generalized exploratory IRT model is employed (De Boeck & Wilson 2004) which is formulated as random intercept model with a logit link and a binomial error (Skrondal & Rabe-Hesketh 2004).

A survey in the republic of Germany was conducted in 2007 to estimate the effects of interest for 5 dichotomous items designed to measure health related quality of live. Each of the possible 120 (5!) orders were administered in a fully balanced design on 960 respondents. Order effects with respect to the item-difficulties were estimated and evaluated with respect to the latent dimension under consideration. Item specific order effects are estimated by means of different contrasts and interactions between the fixed parameters of the model and the appropriate sequence pointer. Estimations were carried out by the module gllamm within STATA 9.2 (StataCorp 2005)

References

- De Boeck P, Wilson M. 2004. Explanatory Item Response Models: A General Linear and Nonlinear Approach. New York, Berlin: Springer.
- Skrondal A, Rabe-Hesketh S. 2004. Generalized Latent Variable Modeling; Multilevel, Longitudinal, and Structural Equation Models. London, New York: Chapman & Hall/CRC.
- StataCorp. 2005. Stata Statistical Software: Release 9. College Station, TX:

How to Estimate and Control for Order Effects Using Generalized IRT Models: Prerequisites and Limits

Rainer Alexandrowicz; University of Klagenfurt, Austria

Herbert Matschinger; University of Leipzig, Germany

Attitude measurement in survey research frequently employs sets of stimuli (items) presented e.g. as questionnaire or a CATI. Either way a certain sequence of these stimuli exists/has to be chosen. The current study focusses upon the question whether and to what extent the very position of one such item in the succession of items might cause DIF or item bias. Possible source of bias could be (amongst many others/just to mention a few) learning or tiring effects or simply the fact that the respondent

is being asked a question he never has been occupied with before. The statistical model employed is a generalized exploratory IRT model (DeBoeck & Wilson, 2004) which can be formulated as random intercept model with a logit link and binomial error (Skrondal & Rabe-Hesketh, 2004). By means of a simulation study we show how a study has to be designed in order to allow for detecting item bias (DIF) due to ordering. Different kinds of ordering effects are considered and the question of necessary sample sizes is being discussed.

Satisfaction Surveys of External Clients As a Tool to Identify and to Implant Areas of Improvement in the Public Health System

Palmira Tejero Cabello; Consorci Sanitat Integral, Spain

Juana Gómez Benito; University of Barcelona, Spain

Jose Luis Padilla Garcia; University of Granada, Spain

Sandra Juncosa; Consorci Sanitat Integral, Spain

Montse Marfín; Consorci Sanitat Integral, Spain

Santiago Nofuentes; Consorci Sanitat Integral, Spain

The present study uses the satisfaction surveys to incorporate areas of improvement in the acute hospitals of the Consorci Sanitari Integral (CSI). CSI is a company of public administration of social and health services of Catalunya whose main aims are people and the commitment with the society, offering services and health attention with quality and looking after the satisfaction of clients. In its strategic planning to get the client's opinion, it has elaborated a questionnaire which studies the satisfaction of its users in the four key processes of its two hospitals: hospitalization, emergencies, external consultations and ambulatory major surgery. For the elaboration of these questionnaires, in the year 2002 it was formed a consent group for each process, composed by 20 professionals of the two centers who received a specific training on methods to evaluate the perceived quality and to design jointly an useful instrument. The work in groups gave as a result the creation of a unique and homologated model of surveys that allowed the evaluation of the satisfaction and the expectations of patients. To identify easily the areas of improvement, the method of the traffic light was used which consists in assigning to each one of the results obtained in each question a smiling face (if the result is positive and it is not necessary to act to improve it) and a angry face (if the result is negative and it must be necessary some action to improve a certain aspect) as they approached or not to the good result defined previously in function of previous studies. The target population : People assisted during the previous year in the key processes. Exclusion approaches: patients entered by Psychiatry, exitus. Selection of the sample: simple aleatory sampling. Sample size: to estimate a prospective proportion of 50% with a level of trust of 95% and a precision of 7%, the minimum of fellows is of 196 in each area. The surveys were carried out phonly by the same interviewer. The answer in Likker scale and dicotomic. Global punctuation for area from 0 to 10. Starting from the results obtained in each process, center and year, there are identified the areas that each hospital must work as an action of improvement, what supposes an important source of information for the strategic planning of the CSI.

Utility of Polytomous Differential Item Functioning Techniques in Cross-cultural Survey Data

María-Dolores Hidalgo ; University of Murcia, Spain

Juana Gómez Benito; University of Barcelona, Spain

The wide use of surveys in social research makes high quality construction relevant. Thus, when cross-cultural surveys studies are using is important examinee the sources of error due to translation/adaptation process. In general, few studies have been conducted on Differential Item Functioning (DIF) analysis in survey data as compared with research in educational test and psychological questionnaires data. Some previous works in this area analyse DIF in item-survey to measure attitudes toward mathematical and science, Family Survey, Economic Expectations and Attitudes, and Attitudes toward

the Environment. These approximations are related to DIF stability over subsamples from a single population, more frequently to assess gender-related DIF and using techniques such as Mantel-Haenszel and Logistic Regression. Moreover, when the items were with more than two levels of response they were dichotomized in order to use dichotomous DIF procedures. However polytomous DIF techniques are more suitable for this item format. The aims of this work are: a) to show the advantages of Polytomous Differential Item Functioning analysis in Cross-Cultural Surveys, and b) to analyse some problems regard to this area.

Investigating Measurement Invariance of the Infant-toddler and Emotional Assessment (ITSEA) Scale Across Demographic Groups in the Netherlands

Francisca Galindo Garre; TNO Quality of Life, Netherlands

Matty H. Crone; TNO Quality of Life, Netherlands

Reinoud D. Stoel; University of Amsterdam, Netherlands

As in most European countries, the Dutch population has become very diverse in the last years with respect to ethnicity. Several instruments are commonly used by health institutions to diagnose for social-emotional/behavioural problems in young children. It is desirable that diagnoses based on these instruments are only affected by the presence of the relevant problems and not by other variables, such as ethnicity and gender.

One of the relevant instruments is the Infant-Toddler and Emotional Assessment (ITSEA). Although psychometrical properties of the American version of the ITSEA have been extensively investigated, no research has yet been done to investigate whether the ITSEA results in unbiased diagnoses with respect to ethnicity and gender. The purpose of this study is, therefore, to evaluate the measurement invariance of the ITSEA. Multi-group confirmatory factor analysis that incorporates between-group mean structure (Meredith, 1993) will be used to test the various invariance hypotheses. We will use data from a sociodemographically diverse birth cohort sample of parents of 14 months old children born in the Netherlands. The implications of the results for the study of gender, ethnic, and cultural differences associated with social-emotional and behavioural problems will be discussed.

Direction of Response Categories, Specificity of Items and Response Behavior

Dagmar Krebs; University of Giessen, Germany

Jürgen H.P. Hoffmeyer-Zlotnik; Centre for Survey Research and Methodology (GESIS-ZUMA), Germany

Based on a split-ballot as well as repeated measurement design, the paper approaches the question if a change in the direction of response categories does affect response behaviour. Respondents were asked to answer different attitude questions by using an 8-point scale offering increasing (split 1) or decreasing (split 2) intensity of how much an item applies to them or how strongly they agree with an item. Only the extreme points of the 8-point scale were labelled. Data collection was done with three different versions of the questionnaire: First all persons got the same questionnaire using a 1-8-point scale where answering categories ranged from “does not at all apply (1)” to “applies completely (8)”, scale points in between had only numbers attached to them. Second, the same individuals were asked again (about three weeks later) but this time they were assigned to different split versions. For one group of respondents again a 1-8-point scale was used but response categories were now labelled as “applies completely (1)” to “does not at all apply (8)” (split 1). The second group got a questionnaire containing the reverse version of the first measurement where response categories were labelled as “applies completely (8)” to “does not at all apply (1)” (split 2). The third group got the same questionnaire as in the first measurement thus building the test-retest version (split 3).

The hypothesis is tested that response behaviour differs depending on the direction of response categories as well as on the values attached to the categories.

However, item content is supposed to be an intervening variable. If items describe rather general states of mind as e. g. “something should be ...”, it is expected that distributions (as well as means)

do not differ depending on the scale used. If, however, items contain specific statements related to respondents themselves as e.g. "It is important for me . . .", respondents might feel a stronger appeal to clarify their position by checking the appropriate response category. Compared are statements about how central family or work should be for a person and items assessing achievement and job motivation. Direction of response categories is expected to produce different response behaviour in the former but not the latter statements.

Key words: response categories, measurement quality, response behaviour.

2.8 Enhancing Survey Methodology with Qualitative Methods

Session coordinated by:

Bojana Lobe; University of Ljubljana, Slovenia

Presentations

1. *Mixing Methods in Evaluation Research: Notes from the Field* – I. Zandberg (United States), C. Robins (United States), S. Berkowitz (United States)
2. *A Delicate Dialectic: Mixing Qualitative and Quantitative Methods in Studies for Us Federal Government Agencies* – S. Berkowitz (United States), C. Robins (United States), I. Zandberg (United States)
3. *Combining Qualitative and Quantitative to Improve Methodology for a Research Project on Homelessness* – M. Quaglia (France)
4. *A Qualitative Post-test on the 4th Ewcs* – G. Vermeylen (Ireland)
5. *Cultural Differences Between Ethnic Minorities: Assessing and Understanding the Impact of Cultural Bias* – M. Morren (Netherlands)
6. *Inaccuracy of Birthday Respondent Selection Methods in Mail and Telephone Surveys* – R. Schnell (Germany), S. Ziniel (United States), E. Coutts (Switzerland)
7. *Deliberative Surveys. an Analysis of Their Design, Execution and Results Following the First Experience in Spain.* – M. Cuesta Azofra (Spain), S. Pasadas Del Amo (Spain)
8. *Concept Changing: 'a Tool' for Looking for New Ideas in Research Work* – S. Seykova (Bulgaria)

Presentations in the session “Enhancing Survey Methodology with Qualitative Methods”

Mixing Methods in Evaluation Research: Notes from the Field

Izabella Zandberg; Westat, United States

Cynthia S. Robins; Westat, United States

Susan G. Berkowitz; Westat, United States

The two central questions posed by evaluation research are whether the evaluated program accomplished its intended outcomes, and why. Usually, particularly in the case of multi-site programs, the focus of the applied research is on the measurable effect of the program, and fewer resources are committed to discover why the variability in its success (or failure) arises across different sites. Consequently, quantitative data are considered to be the primary sources for decision makers, and qualitative approaches are usually used as supplemental means of data collection. Overall, the utilitarian value of qualitative methods in the applied research environment is seen mostly in their ability to provide exploratory insight to inform survey development, or as a means to explore in-depth findings generated through survey research. On rare occasions quantitative research is interwoven with qualitative work into a research design from the early phase of study development in the evaluation inquiry.

This presentation is based on an experience of a qualitative research team working in a large social science research company in the United States. Most of the research at this company is funded by various agencies within the U.S. government. Qualitative methods – focus groups, in-depth interviews, case studies – are often funded with the expressed goal of enriching the quantitative approaches. The presenters will discuss two mixed-methods projects that were conducted under such government contracts. The presentation will describe the political implications of using qualitative methods in applied research, and in particular why the legitimacy of these methods continues to be questioned in the United States.

A Delicate Dialectic: Mixing Qualitative and Quantitative Methods in Studies for Us Federal Government Agencies

Susan G. Berkowitz; Westat, United States

Cynthia S. Robins; Westat, United States

Izabella Zandberg; Westat, United States

In the “benign abstract,” the idea of mixing qualitative and quantitative methods in the same study design has gained increasing acceptance. However, it can still be extremely challenging to come to terms with the realities of what actually happens—and what it means—when divergent sets of epistemological assumptions and different methodological approaches are brought to bear on a single problem or research question. This paper will discuss the particular challenges and rewards of carrying out mixed method applied research studies in a non-academic, contract research environment. Emphasis will be placed on the epistemological and methodological issues that must be addressed (directly or indirectly), as well as how these concerns intersect and interact with the realities of funding and intellectual hegemony (power of definition).

The authors will provide case examples from 4 studies for US federal government agencies that achieved varying degrees of success in “making a good marriage” between the qualitative and quantitative (survey) portions of their respective designs. Qualitative methods in question included ethnography, focus groups and in-depth interviewing. The paper will identify factors that helped as well as those that hindered the development and maintenance of a strong, positive, mutually informing relationship between the study components, and will advance an initial typology of “successful marriages” and “less successful “marriages in this realm that may be of wider relevance and application.

Combining Qualitative and Quantitative to Improve Methodology for a Research Project on Homelessness

Martine Quaglia; Institut National d'Etudes Démographiques, France

As a member of the Survey department of the National Institute for Demographic Studies (INED) in France, I have taken part over the past ten years in a number of studies related to homelessness, working with teams differing both in approach and background including demographers, statisticians, psychiatrists (clinicians and epidemiologists), sociologists, anthropologists and psychologists. These experiences led me to question the legitimacy of the still discernible rivalry between supporters of the qualitative and supporters of the quantitative which is no longer adapted to the reality of the research.

Combining both methods is an obvious advantage for the quality of data thus produced. Because the two methods don't ask the same questions and therefore, produce answers on different themes, because the combination, when it is done, enlarges the researcher point of view and enriches the method and tools for a better analysis of situations.

Through a brief description of some of the methodological issues raised during a research programme for which both methods were used as complementary, we will try to discuss some answers that can be provided through the association of the two methods.

Key words : combining qualitative and quantitative methods, survey methodology

A Qualitative Post-test on the 4th Ewcs

Greet Vermeylen; European Foundation for Living and Working Conditions, Ireland

The 4th European Working Condition Survey (4th EWCS) was conducted in Autumn 2005 in 31 countries by European Foundation for Improvement of Living and Working conditions. As one of extensive quality control measures, a qualitative post-test on the survey was carried out for the first time, thus combining qualitative methods with quantitative survey research. The main objectives were:

- a quality control measure assessing the quality of the 4th EWCS questionnaire,
- a tool for interpretation of survey results
- a tool for developing new questions
- to help to understand the role of national infrastructure, cultural context and gender in shaping respondents' answers
- to gather more information on the subject of development at work and employability.

The qualitative post-test was conducted in five countries representing different European realities: Austria, Czech Republic, Finland, Portugal and the UK. In each of the countries, 20 interviews were conducted among the respondents of 4th EWCS, representing women and men in different age groups, occupations, sectors and educational levels. The interviews were implemented as a combination of structured survey questions and qualitative interviewing approach: the interviewees were presented a selected number of questions from the original questionnaire and they were encouraged to contemplate to what kind of situations they referred to by their answers and how they found the questions and responses alternatives. The additional open-ended questions on content of the job, work organisation, training and other ways of self-development sought to map the phenomena of development at work and employability and their role for the interviewees.

National analysis was done at two levels: first, assessing the technical quality of the questionnaire and secondly, analysing such themes around the phenomena of development at work and employability, emerging from the interviews. At the second stage, the consolidated report drew together the main findings of the national reports, looking for common nominators in this respect and highlighting national differences where they emerged.

Given the challenges of cross-cultural survey in general, and the wide scope of different countries the EWCS covers in particular, it seems that the selected questions worked out pretty well in the five post-test countries representing different working conditions. However, in few cases it became apparent

that some specific national results did not quite reflect the intended meaning and were not comparable with the other countries.

In the technical part, we examined every question individually. We looked at possible translation shortcomings, but also at the variety of ways the questions were interpreted by different groups of informants, reflecting differences in their personal, job-related and cultural contexts. The phenomena part looked at the links between different themes and eventual shortcomings in the survey. The current questionnaire covers a variety of (basic) issues with regard to development at work, such as opportunities to learn new things and grow at work, training participation and various features of work which can be considered developmental. However, the qualitative approach revealed that there was more into these various themes than the current survey questions were able to grasp. This could allow development of further questions in the next edition of the EWCS on this theme.

Cultural Differences Between Ethnic Minorities: Assessing and Understanding the Impact of Cultural Bias

Meike Morren; University of Tilburg, Netherlands

Many national surveys are held among culturally different populations. Often it is assumed that the responses are comparable, although it is uncertain whether the respondents of different cultural backgrounds understand the questions in a similar way. To discover to what extent differences in answers between cultural groups reflect different opinions or different response styles is difficult.

Respondents can, due to their cultural background, differ in their understanding of the intention of a question, or their interpretation of the concepts used in the question or the meaning they attach to the question. The problematic issue in cross-cultural research is that differences in response styles can relate to many respondent's characteristics such as age, educational level or cultural background, but also to characteristics of the interviewer or the question.

We argue that a combination between research methods could contribute to the distinction between culture and response style. On the one hand, differences in response styles will be detected by use of statistical models; on the other hand the results from this analysis will be further investigated by means of focus groups and cognitive interviewing. By first detecting differences among culturally diverse respondents in a national survey, a general impression is gained on the differences that are present. Focus groups are particularly useful to give insight whether the differences in responses are caused by a different idea about the subject of the question. The method of cognitive interviewing can be used to discover the ways in which respondents from different cultural backgrounds approach a question differently. Thus, by combining these methods, a more encompassing understanding of sources of cultural bias in national surveys including ethnic minorities can be achieved.

Here, the advantages and the pitfalls of the quantitative and qualitative methods to detect cultural differences in national surveys are discussed. The study contributes to the existing knowledge of differences in response effects between cultural groups, because 1) cultural groups are selected within one country, which enables us to bring together and compare different cultural groups within one research setting, 2) a mixed-methods research design is applied.

Inaccuracy of Birthday Respondent Selection Methods in Mail and Telephone Surveys

Rainer Schnell; University of Konstanz, Germany

Sonja Ziniel; University of Michigan, United States

Elisabeth Couffts; ETH, Switzerland

If households are used as sampling units in surveys of people, a method for random selection within households is required. Birthday methods are common in telephone and mail surveys. Here, we report the results of two studies evaluating such selection procedures. First, we compare a last-birthday selection with the selection implied by a self-reported household roster in a German national telephone survey (n=1,615). About 12% of the selections within multi-adult households appeared erroneous.

Second, we compare the selection within households using the last-birthday method in a mail survey with individual-level data from an official German town register (n=1,069). About 33% of selections within multi-adult households appeared to be in error. Based on a qualitative study of households within which selection errors took place (n=30), we conclude that two age-cohort specific intra-household mechanisms are responsible such errors. These results were validated by logistic regressions within subgroups of the CATI-survey.

Deliberative Surveys. an Analysis of Their Design, Execution and Results Following the First Experience in Spain.

María Cuesta Azofra ; Centro de Investigaciones Sociológicas (CIS), Spain

Sara Pasadas Del Amo; Institute for Social Studies of Andalusia, Spain

Deliberative surveys emerged in the mid-nineties as a novel tool designed to measure public opinion. The notion behind this type of survey is quite simple. After conducting an initial survey of a representative sample of the study population, respondents are invited to take part in an encounter known as a “deliberative forum”. In this forum, the participants receive balanced information about the issue at hand and engage in a process of citizen deliberation in small, randomly selected groups. Upon conclusion of the forum, participants are again asked to fill out the original survey to determine if exposure to the information and the deliberative process lead to changes in the respondents’ attitudes and opinions regarding the issue.

Following twenty experiences of this kind around the world, a group of researchers from the Institute for Advanced Social Studies of Andalusia (IESA) and the Centre for Sociological Research (CIS) conducted the first deliberative survey in Spain in the city of Cordoba. The topic of the survey was “Night-time Youth Leisure.” The aim of our paper is to present the initial results of this experience, paying special attention to the methodological design, the different problems that arise when conducting the survey, the added value that can be gained from this methodological tool as regards research results and the usefulness of the results obtained in this deliberative survey.

Concept Changing: 'a Tool' for Looking for New Ideas in Research Work

Svetlana Seykova; Bulgarian Academy of Sciences, Bulgaria

This paper presents the author’s vision for using the “concept changing” approach as a specific mental “tool” for stimulating researchers’ innovative thinking.

New concepts (or modifications of existing concepts) are ways to distance ourselves from the familiar theories and experience in order to look for new ideas and more effective solutions. Paradigm shift (by T. Kuhn) is a classic example of how the “concept changing” approach stimulates the creation of new ideas in the area of research knowledge. This approach proves equally fruitful in finding innovative solutions in the area of business, or the activities of social, political and other institutions/organizations. The foundation of the “concept changing” approach is E. De Bono’s theory about the asymmetry of creative thinking. In this case, it is viewed as a “tool” for stimulating the creation of new ideas and solutions particularly in social science research.

The author introduces her definition of the term “concept”, specific to the search of new ideas. Following E. De Bono’s ideas, the author briefly presents the basics of the asymmetric thinking process and the barriers it poses. She reviews the primary sources of ideas for concept changing, as well as a few impressive examples of the effects of the presented approach in research studies. She also examines the key advantages of the approach in looking for more effective ideas and solutions in both science and life.

Key words: concept approach in innovative thinking; obstacles caused by asymmetric thinking process; creative and lateral thinking; concept approach in research work

3 Measures for Political Science Research

Sessions in Measures for Political Science Research

Analysing Attitudes Towards Migration with Large Comparative Survey Data	76
Measuring the Quality of Democracy on the Basis of Survey Data	79
Human Values	83
Measuring and Explaining Trust	87
European Values Study	93

3.1 Analysing Attitudes Towards Migration with Large Comparative Survey Data

Enric Martínez-Herrera; Centro de Estudios Políticos y Constitucionales (CEPC), Spain

Presentations

1. *Racial Prejudice and Opposition to Anti-racist Policies in Europe: Individual and Contextual Predictors* – J. Vala (Portugal), A. Ramos (Portugal), C. Pereira (Brazil)
2. *The Relation Between Human Values and Attitudes Toward Immigration Policy - a Comparison Across 19 European Countries* – B. Meuleman (Belgium), E. Davidov (Germany), P. Schmidt (Germany), J. Billiet (Belgium)
3. *Attitude Toward Foreigners: Reliability and Validity of Unipolar Versus Bipolar Scaling* – D. Krebs (Germany), J. Hoffmeyer-Zlotnik (Germany)

Presentations in the session "Analysing Attitudes Towards Migration with Large Comparative Survey Data "

Racial Prejudice and Opposition to Anti-racist Policies in Europe: Individual and Contextual Predictors

Jorge Vala; University of Lisboa - ICS, Portugal

Alice Ramos; University of Lisboa - ICS, Portugal

Cícero Pereira; Universidade Católica de Goiás, Brazil

This paper analyses the predictors of prejudice and opposition towards anti-racist policies in Europe, using data from European Social Survey 2002. Two different levels of predictors are considered: individual level, and contextual level. At the individual level threat perception and social values were selected as predictors, and at the contextual level we considered immigration policies, economic factors, migration and political and ideological factors. The questions addressed are: 1) what are the predictors of prejudice at both individual and contextual level?; 2) what is the impact of prejudice on opposition towards anti-racist policies?; 3) is there a mediating role of threat perception on the relation between prejudice and opposition to anti-racist policies? Results show that 1) differences between countries concerning racial prejudice and opposition to anti-racist policies are mainly explained by individual variables, being share vote in extreme right the only contextual variable that showed some predictive power, but only on opposition to anti-racist policies; 2) at the individual level, the adherence to the values of self-transcendence and conservation are the best predictors of prejudice; 3) Concerning opposition to anti-racist policies, the best predictors are prejudice, perception of symbolic threat and adherence to self-transcendence values; 4) finally, although there is a direct effect of racial prejudice on opposition to anti-racist policies, this relationship is partially mediated by perceptions of threat, mainly by symbolic threat.

The Relation Between Human Values and Attitudes Toward Immigration Policy - a Comparison Across 19 European Countries

Bart Meuleman; University of Leuven, Belgium

Eldad Davidov; University of Köln, Germany

Peter Schmidt; University of Giessen, Germany

Jaak Billiet; University of Leuven, Belgium

The literature contains numerous studies that attempt to gain deeper understanding of anti-immigration feelings and negative out-group attitudes. Traditionally, these studies have focused on the effects of structural variables, economic interests, situational triggers, perceived threat and symbolic motives. Far fewer studies investigate the direct relation between human values and out-group attitudes. Nevertheless, the theoretical relevance of values to explain out-group attitudes is paramount. After all, human values can be seen as abstract principles that underpin attitudes towards more specific objects, such as immigration or out-groups (Rokeach 1973).

In this paper, we estimate and compare the effects of two value dimensions taken from the theory of basic human values (Schwartz 1992) – namely self-transcendence and conservation- on attitudes toward immigration policies in 19 countries. Data from the first round (2002-03) of the European Social Survey (ESS) is utilized. This cross-national survey measures basic human values with a new 21-item instrument. Attitudes toward immigration are operationalized using two dimensions: willingness to allow immigrants into the country, and support for imposing conditions on immigration.

Effects of the value dimensions on immigration attitudes are compared across 19 nations using multiple-group multiple-indicators structural equation modelling (MGSEM). The critical statistical legitimacy for comparing these effects across countries is discussed in detail, and measurement invariance

is evidenced (Steenkamp & Baumgartner 1998). The MGSEM provides strong support for our hypotheses in 17 countries: self-transcendence displays a positive effect on support for immigration, and conservation a negative effect. It is shown that among the countries in the study, clusters of countries with equal effect sizes can be distinguished.

References

- Rokeach, M. (1973). *The nature of human values*. New York: The Free Press.
- Schwartz, S. H. (1992). Universals in the content and structure of values: theoretical advances and empirical tests in 20 countries. P. Zanna (ed.), *Advances in experimental social psychology* (Vol. 25pp. 1-65). San Diego: Academic Press.
- Steenkamp, J-B & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research, *Journal of consumer research*, 25, 78-90.

Attitude Toward Foreigners: Reliability and Validity of Unipolar Versus Bipolar Scaling

Dagmar Krebs; University of Giessen, Germany

Jürgen H.P. Hoffmeyer-Zlotnik; Centre for Survey Research and Methodology (GESIS-ZUMA), Germany

Should attitudes be measured by a unipolar or by a bipolar scale? In attitude research it is often strongly recommended to view attitudes as bipolar dispositions of favorability or unfavorability toward an attitude object.

With respect to attitudes toward foreigners, the question of measurement quality (in terms of reliability and validity) of either scale is approached by using a combination of split ballot design and repeated measurement. While the split ballot design gives information about different means, standard deviations as well as correlations of responses to either the unipolar or the bipolar scale, repeated measurement aims at introducing method effects affecting reliability and/or validity of measurement using either the unipolar or the bipolar scale. This kind of reliability and validity assessment is only possible if repeated measures of the same constructs are available.

Data on attitude toward foreigners were collected by using a 1-7 point scale with verbalized extremes of “not at all agree (1)” and “completely agree (7)” for the unipolar scale. The bipolar scale contained also seven points verbalized at the extremes as “not at all agree (-3)” and “completely agree (+3)”. Between the extremes response scales contained numbers of 2,3,4,5,6 for the unipolar or -2, -1, 0 1, 2, respectively for the bipolar scale. Respondents in both versions were asked to respond by writing the numbers into the questionnaire (since it was a paper and pencil design).

In addition to assess differential response behaviour depending on the uni- versus bipolar character of response scales, the effect of question wording (positively versus negatively formulated items for assessing the attitude toward foreigners) can be tested under the two scaling conditions. It can also be shown that correlations between negatively and positively formulated items differ depending on the uni- or bipolar response scale used.

Results revealing different distributions of either scale type and showing that means are higher, standard deviations lower and correlations stronger on the bipolar compared to the unipolar scale will be presented as well as comparison of measurement models with respect to factorial structure and latent means. Additionally, effects of either method on reliability and validity can be demonstrated.

Key words: response behaviour, response scales, uni- versus bipolar scaling, attitude toward foreigners

3.2 Measuring the Quality of Democracy on the Basis of Survey Data

Session coordinated by:

Hanspeter Kriesi; University of Zürich, Switzerland

Marc Bühlmann; University of Zürich, Switzerland

Presentations

1. *Measuring Democracy: Possibilities and Limits of Citizen's Assessments.* – H. Lauth (Germany)
2. *The Quality of Participation and Its Institutional Determinants: an International Comparison* – M. Bühlmann (Switzerland), H. Kriesi (Switzerland)
3. *Measuring a Public's Fitness for Democracy: Individual-level and Aggregate-level Findings.* – C. Welzel (Germany)
4. *Social Capital and Political Participation in Britain: is There a Relationship at the Individual Level?* – K. Yang (United Kingdom), W. van der Veld (Netherlands)
5. *The Quality of the ESS Measures for Political Efficacy* – W. Saris (Spain)
6. *Mapping Attitudes Towards Democracy in Europe. a Framework for a Future Research Agenda.* – A. Trechsel (Italy)

Presentations in the session "Measuring the Quality of Democracy on the Basis of Survey Data"

Measuring Democracy: Possibilities and Limits of Citizen's Assessments.

Hans-Joachim Lauth; FernUniversität Hagen, Germany

The paper discusses the possibilities of using surveys in measuring democracy. Starting with an overview and analysis of the current assessments, the central aim of the paper treats the question whether surveys can be a promising avenue in the assessment of democracy. In the literature on this topic one can identify reasons pro and contra using surveys. The position developed in my contribution will argue that the appropriate question is not "Using - yes or no" but which aspects that are measured by surveys can contribute to democratic assessment and which not. For example: should "overall satisfaction with democratic performance" be a useful indicator? Opening this debate the paper will identify these aspects of survey which are useful or can be useful and those aspects which should be treated with cautiousness or should be ignored anyway. Finally the consequences for further research are treated.

The Quality of Participation and Its Institutional Determinants: an International Comparison

Marc Bühlmann; University of Zürich, Switzerland

Hanspeter Kriesi; University of Zürich, Switzerland

Advocates of participatory democracy highlight the importance of participation (Barber 1984; Pateman 1970). Greater access to politics as well as more direct say for the people are not only seen as possibilities for self-transformation (Warren 1992) but also as a remedy for the increase of disillusion with politics (Norris 1999). Critics of this view advance the dangers of populism: More direct participation by the people finally undermines the important role of intermediary organisations and lead to a tyranny of a unitary majority which is controlled by charismatic leaders (Abts and Rummens 2007; Mény and Surel 2000). One counter-argument against this view stems from the radical model of democracy which foremost focus on the quality of participation (Cohen 2004; Offe and Preuss 1991; Warren 1996). In this view, qualified participation means argument-based participation by informed and interested people which resist to the seduction of populists. Additionally, the more individuals have the opportunity to participate the higher the level of information and of political interest.

Institutions that broaden the access to politics, allow for more direct participation, and facilitate participation should therefore enhance the quality of participation.

We investigate this approach by conducting a cross-country comparison. First, we propose some simple measures of the 'quality of participation'. Second, using survey data, we try to assess the quality of participation for a number of countries. Third, we test the influence of different institutional settings on our quality measures.

References

- Abts, Koen, and Stefan Rummens. 2007. *Populism versus Democracy*. Political Studies fc.
- Barber, Benjamin R. 1984. *Strong Democracy. Participatory Politics for a New Age*. Berkeley: University Press.
- Cohen, Joshua. 2004. *Radical Democracy*. *Swiss Political Science Review* 10 (4):169-180.
- Mény, Yves, and Yves Surel. 2000. *Par le peuple, pour le peuple. Le populisme et les démocraties*. Paris: Fayard.
- Norris, Pippa, ed. 1999. *Critical Citizens: Global Support for Democratic Government*. Oxford: Oxford University Press.
- Offe, Claus, and Ulrich K. Preuss. 1991. *Democratic Institutions and Moral Resources*. In *Political Theory Today*, edited by D. Held. Cambridge: Cambridge University Press.
- Pateman, Carole. 1970. *Participation and Democratic Theory*. London.

Warren, Mark. 1992. Democratic Theory and Self-Transformation. *American Political Science Review* 86:8-23.

Warren, Mark E. 1996. What Should We Expect from More Democracy?: Radically Democratic Responses to Politics. *Political Theory* 24 (2):241-270.

Measuring a Public's Fitness for Democracy: Individual-level and Aggregate-level Findings.

Chris Welzel; Bremen International University, Germany

Survey research routinely assumes that the best way to measure how mature people are to practice democracy is to ask them directly about political preferences and about attitudes towards aspects of democracy. Based on evidence from the World Values Surveys this contribution argues against this premiss, showing that more basic attitudes towards life and people are much more indicative of both a "democratic personality" (at the individual level) and a "democratic culture" (at the aggregate level) than standard indicators of support for democracy. In particular an index of emancipative values turns out to be an astoundingly strong correlate of various objective measures of democracy.

Social Capital and Political Participation in Britain: is There a Relationship at the Individual Level?

Keming Yang; University of Reading, United Kingdom

William van der Veld; University of Amsterdam, Netherlands

Since Putnam's well-known but controversial study on the relationship between social capital and democracy in Italy, there have been two shifts of analytical focus in the literature. The first is substantive (from democracy to political trust) and the second is analytical (from individual level to aggregate levels). A widely accepted argument is that social capital works most effectively for promoting democracy by increasing the level of political trust at aggregate levels. The first part of this paper is consisted of two counter-arguments: (1) Political trust has little to do with social capital and democracy; (2) Social capital can be an individual property and thus can be linked to political participation at individual level. Therefore, it is more meaningful to specify the targeted relationship as a direct one between social capital and political participation. Empirically, I test these relations at the individual level by analyzing the UK data from European Social Survey (ESS). The study is the first step of a larger project that ultimately aims to measure the relationship between social capital and political participation in all participating countries of ESS.

The Quality of the ESS Measures for Political Efficacy

Willem E. Saris; ESADE, Universitat Ramon Llull, Spain

Through the years many different approaches to measure political efficacy have been proposed. Alternatives were proposed based on different numbers of items, alternative items, alternative measurement models and alternative response scales. After an elaborate discussion the European Social Survey has chosen its approach to measure political efficacy based on earlier work of Angelika Vetter. In this paper we discuss the quality of the ESS measures with respect to reliability and validity of the single items and the composite scores. It is also indicated how composite scores can be calculated and in how far these measures are comparable across countries.

Mapping Attitudes Towards Democracy in Europe. a Framework for a Future Research Agenda.

Alexander Trechsel; European University Institute, Italy

The nascent research agenda presented is one of the rather large goals of the newly created 'European Union Democracy Observatory' (EUDO) located at the European University Institute's Robert Schuman Centre for Advanced Studies. The research agenda mirrors the conceptualization of research clusters identified within the EUDO framework.

There is little novel in the fact that European democracies are currently facing a prolonged period of self-examination and both a challenging internal and external context. However, there is a palpable anxiety regarding democratic deficits at various levels of government compounded by major changes in representation. In this context, therefore, the proposed research agenda seeks to learn more about democracy's most important actors, its citizens, by examining their attitudes towards democracy at the various levels of the EU.

Existing scholarship and data in this field appear too limited, either because they focus largely on transitional polities or the elusive concept of satisfaction with democracy. Furthermore, comparative, EU-centred approaches that examine public attitudes within and across the national level, as well as above, are singularly lacking. In addition, little attempt has been made to conceptualise the multiple facets of democracy that help structure citizens' opinions. Public opinion scholarship strongly stresses the role of elites and the media in opinion formation. However, the interaction regarding attitudes towards democracy between citizens, elites and the media in Europe is uncharted territory.

Hence this agenda aims to overcome this glaring gap by proposing the design of a large-scale public opinion survey to be carried out in all EU member states. This would comprise batteries of questions designed to measure attitudes towards democracy. The questions of the survey will be submitted simultaneously to elites and media actors at the national and European level. The analysis of this data would permit unique comparisons between citizen, media and elite attitudes towards the multiple dimensions of democracy at all levels of the European polity. Results from such an analysis would prove valuable to a wide spectrum of social scientists, policy-makers and civil-society actors, consequently generating a range of further potential studies.

3.3 Human Values

Session coordinated by:

Eldad Davidov; University of Köln, Germany

Presentations

1. *Predicting Opposition Towards Immigration: Economic Resources, Social Resources and Moral Principles* – A. Ramos (Portugal), J. Vala (Portugal)
2. *Measuring Value Orientations Across Nations: Analysis of the Quality of the European Social Survey for Measuring Values* – D. Knoppen (Spain)
3. *Values and Social Classes in Europe* – J. Ferreira de Almeida (Portugal), R. Brites (Portugal)
4. *New Meanings of the Family on the Swing with Work: an European Overview* – A. Torres (Portugal), R. Brites (Portugal), B. Coelho (Portugal), I. Cardoso (Portugal)
5. *Political Support. Explanations in Comparative Perspective in Europe* – E. Bartolomé Peral (Spain)
6. *Exploring Social Capital from a Cross-paradigmatic Point of View* – F. Heselmans (Belgium)

Presentations in the session "Human Values"

Predicting Opposition Towards Immigration: Economic Resources, Social Resources and Moral Principles

Alice Ramos; University of Lisboa - ICS, Portugal

Jorge Vala; University of Lisboa - ICS, Portugal

This study analyses the predictors of opposition towards immigrants from 'different ethnic groups' and 'poor countries' in 5 European countries (Portugal, Germany, Netherlands, France and United Kingdom), using data from the ESS 2002. Besides Portugal, a country that moved from being a country of net emigration to being a new host country for immigrants, the other ones were selected according to their main policies of immigrants' integration. Opposition towards immigration (OTI) is analysed using three theoretical models: a) the economic self-interest model that proposes that opposition towards immigration may be due to economic factors; b) the social capital model according to which social trust and self-reliance on political and social system may shape peoples' opinions on the benefits of immigration; c) the Schwartz's human values model, based on which it is possible to predict that some values facilitate OTI, whereas others facilitate openness to immigration. The hypotheses tested are: a) there is a negative correlation between economic well-being and OTI; b) a negative correlation between social capital and OTI; c) a positive correlation between both conservation and self-enhancement values and OTI and a negative correlation between both self-transcendence and openness to change values and OTI; d) Moreover, social values model will predict opposition towards immigration over and above the other models. Results globally support the formulated hypotheses.

Measuring Value Orientations Across Nations: Analysis of the Quality of the European Social Survey for Measuring Values

Desiree Knoppen; ESADE, Universitat Ramon Llull, Spain

Scientific and public discourse across a broad range of social science disciplines has discovered the value construct already some time ago. The value construct is also an important constituent of the European Social Survey (ESS) that aims to develop and conduct a systematic study of changing values, attitudes, attributes and behavior patterns within European polities. The employed values model and method of operationalization in the ESS stems from Shalom Schwartz. Whereas the general Schwartz model is supported by several studies, the abbreviated instrument for the ESS counts only with a few critical subsequent analyses. Therefore, the overall aim of the present paper is to determine the quality of the ESS instrument for measuring values and to subsequently analyze possibilities for its improvement. This is done in several steps. First, the operationalization of the different values is studied and conclusions regarding the substantial character of the model are drawn. Then, the quality of each of the items is determined by using the SQP (survey quality prediction) - program, for the English, Dutch and German versions of the survey. The possibility to compare these three countries is analyzed, based on tests of configural, metric and scalar invariance of a first order structural equation model that includes measurement error. After that, the optimal weights for computing the composite scores are derived and the quality of the measure is presented and analyzed. Finally, the external validity of the values construct is evaluated by relating it to other relevant constructs of the ESS such as political participation and demographic characteristics.

Values and Social Classes in Europe

João Ferreira de Almeida; Instituto Superior de Ciências do Trabalho e da Empresa - ISCTE, Portugal

Rui Brites; Instituto Superior de Ciências do Trabalho e da Empresa - ISCTE, Portugal

The analysis of values currently assumes a central position in social research, which tends to conceptualize it as “organized and relatively durable systems of preferences”. Values are expressed in a certain culture, understanding culture in its anthropological meaning, as an “extended and diversified aggregate of characteristics which, on the limit, is synonym of the notion of society”. (Almeida, 1994)

It is in this context that the sociological perspective analyses the significance of values in itself and as behaviour inductors. At the same time it seeks its roots in social classes, in groups and in individuals, without neglecting big aggregates constituted by national States and their eventual diversified identities. As referred by Schwartz, who has proposed the basis for a structural theory of values, these express “motivational aims and they differ, precisely, according to the aims expressed” (Schwartz, 1996).

In this paper we intend to show how class structures strongly influence the patterns of national values, through the analysis of values collected in 23 European countries, in ESS (European Social Survey) round 2, 2004. Those will be related to social classes, in a typology composed by six categories: employers and executives; private professionals; professionals and managers; self-employed workers; routine employees and industrial workers.

New Meanings of the Family on the Swing with Work: an European Overview

Analia Torres; Instituto Superior de Ciências do Trabalho e da Empresa - ISCTE, Portugal

Rui Brites; Instituto Superior de Ciências do Trabalho e da Empresa - ISCTE, Portugal

Bernardo Coelho; Instituto Superior de Ciências do Trabalho e da Empresa - ISCTE, Portugal

Inês Cardoso; Instituto Superior de Ciências do Trabalho e da Empresa - ISCTE, Portugal

Based on the European Social Survey 2004 and 2002 data, and also on other European surveys, we analyse and discuss the realities of work and family in Europe assessing gender differences and similarities in a cross country perspective, using different sorts of multivariate analysis. A methodological discussion around the effects of using different kinds of questions, and thus obtaining for the same themes different answers, will also be developed.

To understand work and family we must perceive these realities as two fundamental dimensions in a process of constitution of personal life projects. New meanings of the family are emerging for both men and women and for all countries in Europe: a person’s family should be main priority in life and men shouldn’t take as much responsibility as women for home and children. These new meanings are rooted in broader processes of social change, on one hand, on the emergence of a new scenario of gender relations in the family; and at the same time, on the individualization process and pursuit of personal happiness.

But these new meanings of the family are on the swing with work, producing a tensional relation that must be solved in everyday life. Women, like many men, want to commit their lives both to family and to work and work is today a fundamental dimension of women’s social identity. However, differences of resources, institutional, ideological and cultural constraints put them in a position where they find it difficult to reconcile these two dimensions of their lives. Contrary to some essentialist positions, which consider that different types of women choose freely different types of work and care arrangements, we will try to show that women tend to prefer investing both in family and on work.

Political Support. Explanations in Comparative Perspective in Europe

Edurne Bartolomé Peral; University of Deusto, Spain

This proposal focuses on political values and orientations of the European citizenry towards their political systems, paying special attention to the concept of political support. This concept, created by David Easton in 1965 tries to measure these orientations of the individuals to their political communities, regime and authorities. These orientations are susceptible to vary depending on the political object (political community, regime, authorities), and also could substantially vary among societies.

The aim of this proposal would be to present, on the one hand, comparisons among European countries regarding to the levels of political support, according to the different objects of support. On

he other hand, it tries to explain the variations in the effect of the different correlates of political support by country. Political support may vary depending of the effect of different correlates which also vary in their explanatory capacity depending on the society we are looking at. In order to explain how citizenry articulate political support and how this levels of support to the different political objects are explained by countries, two main explanations will be tested; the explanation of socialization and cultural aspects, and, on the other hand, the short-term explanation of the evaluation of performance, specially economic, as main explanations for the levels and nature of political support. For the analysis I'll be using the data of the European Social Survey.

Exploring Social Capital from a Cross-paradigmatic Point of View

Frederic Heselmans; University of Liege, Belgium

Social capital and values belong to different paradigms while being closely linked. This is what we will establish within the framework of analyses relating to a vast quantitative survey carried out in Belgium. On the one hand, the combinations value-identities are linked to socio-politic patterns and various democratic models. In addition, social capital and values converge to reflect the various aspects of social integration.

Statistically, most interesting of all this is that these convergences do not mean that the indicators resulting from various paradigms cancel the ones the others or reflect in a different way same reality. Our attempts at modelling show that the explanatory weight of the ones does not absorb that of the others.

From an epistemological point of view, we think that there is no a priori indication of explanatory precedence of one of these components of the society on the others. Do the values forge the identities and condition the social capital? Are the identities the keystone which supports the whole or is this rather the social capital? It acts there like debates which led some "to choose their camp" and to privilege an approach rather than the other... We will try to go up that it is to better privilege the opening, Feyerabend rather than Popper.

3.4 Measuring and Explaining Trust

Session coordinated by:

Patrick Sturgis; University of Surrey, United Kingdom

Presentations

1. *Can We Trust the General Trust Question?* – P. Sturgis (United Kingdom), N. Allum (United Kingdom), P. Smith (United Kingdom)
2. *The Consistency of Citizen and Municipal Level Indicators of Social Capital and Local Government Performance Indicators* – J. Kampen (Belgium)
3. *Trust, Membership, and Social Networks: a Multi-indicator Multilevel Analysis* – N. Letki (Poland)
4. *Social Trust, Political Trust and Civic Engagement* – N. Allum (United Kingdom), P. Sturgis (United Kingdom), R. Patulny (United Kingdom)
5. *The Cross-cultural Measurement Equivalence of Generalised Trust in the ESS* – T. Reeskens (Belgium), M. Hooghe (Belgium)
6. *Testing of a Model of Swift Trust Formation in Temporary Work Groups* – C. Popa (United States)
7. *Analyzing Public Trust in Social Institutions: the Pooled Ordinal Ratings Approach* – T. Tam (Taiwan)
8. *Dimensions of Personal Trust and Their Origins* – H. Iglič (Slovenia)
9. *Determinants of Trust in Low Income, Minority Neighborhoods in Texas* – L. Franzini (United States)
10. *Social Capital and Political Cohesion: How Exceptional is America?* – E. Harrisson (United Kingdom), M. Howard (United States)
11. *Community Networks Membership, Trust, and Community Engagement: the Case of Local Community Mailing Lists* – I. Talmud (Israel), G. Mesch (Israel)

Presentations in the session "Measuring and Explaining Trust"

Can We Trust the General Trust Question?

Patrick Sturgis; University of Surrey, United Kingdom

Nick Allum; University of Surrey, United Kingdom

Patten Smith; University of Surrey, United Kingdom

The extent to which citizens trust one another is a key concern around the world. Robert Putnam's influential theory of social capital has led to academic and policy interest in comparative and longitudinal variation in the extent to which people feel that they can trust their fellow citizens. In the United States, Britain and other advanced democracies, the long-term trend appears to be downward, leading to pessimistic conclusions regarding the social atomisation and fragmentation of modern societies.

The vast majority of the empirical evidence for this decline in social trust comes from what we refer to as the General Trust Question (GTQ). This question, which has been fielded in a great many national and international surveys, asks respondents to choose whether they think 'most people can be trusted', or 'you can't be too careful'. Clearly, this question is not without its problems methodologically. First, it is worded in a very general way. There is good evidence to suggest that questions of a very general nature tend to be interpreted in diverse ways by respondents and are particularly sensitive to context and question ordering effects (Bishop 2005). Second, in some surveys the GTQ is offered with two response alternatives ('can be trusted' or 'can't be too careful') and in others an 'it depends' option is offered. It is not at present clear what the effect is of offering this additional response alternative and whether it is potential 'trusters' or 'non-trusters' from the binary alternative, who are more likely to choose it.

In this paper we present results from a split ballot experiment to evaluate the sensitivity of the GTQ to these varying response formats and to context and framing effects arising from question ordering. Results indicate that sometimes large differences in levels of trust can emerge as a function of response format and question order.

The Consistency of Citizen and Municipal Level Indicators of Social Capital and Local Government Performance Indicators

Jarl K. Kampen; Vrije Universiteit Brussel, Belgium

We address the problem that conclusions reached at aggregate (e.g., municipalities, regions, states) and micro level (e.g., citizens) in correlational studies of social capital and good governance are incompatible. Regardless of the tautologies in Putnam's reasoning, his conclusions on the effects of social capital at aggregate level (the southern and northern regions of Italy) are not replicated in studies at micro level (e.g., Pennants 2005 study on diversity, trust and community participation in England) or international level (e.g., Knack & Keefer's 1997 study on memberships in formal groups and economic performance). This places doubts on the cross level consistency of measures of social capital. In this paper, we try to shed some light on the issue of operationalization of fuzzy concepts by comparing individual, citizen or survey based measures with municipal or aggregate level based measures of indicators of social capital and good governance. With knowledge about the behavior of indicators of social capital and good governance at micro and aggregate level, we are better equipped to judge the value of studies carried out in isolated contexts, that is, with excess to data on either, but not both, levels. Ultimately, the aim of this study is to answer the question whether conclusions reached by helicopter view correspond to conclusions reached from a frog's perspective. We do this by comparing data collected by Statistics Belgium on the 308 Flemish municipalities, with data collected at citizen level by means of a face-to-face survey in a two stage random sample of 1,248 Flemish inhabitants nested within 80 municipalities. This means that the results apply to the subset of 80 municipalities, for which both micro and aggregate level data are available. However, these municipalities are representative of the total of Flemish municipalities because they constitute a random sample.

Keywords: local government, good governance, social capital, consistency, random effects model

Trust, Membership, and Social Networks: a Multi-indicator Multilevel Analysis

Natalia Letki; Collegium Civitas, Poland

Empirical research on social trust and its covariates tends to focus on the formal forms of involvement (group membership) at the expense of informal participation (networks of relationships with friends and relatives). It also frequently relies on one-item indicators of trust, and usually analyses either individual or aggregate-level relationships. Therefore, the results are likely to underestimate the specificity of social trust formation in various contexts (e.g. established democracies vs. post-Communist states), socio-economic elitism of certain types of activism (e.g. membership in organisations), and the effect of context on individual-level attitudes and behaviour.

This paper will use a 2001 ISSP Social Networks Survey to create multi-item measures of various dimensions of social capital, such as interpersonal trust, membership in associations, and informal social networks, and analyse relationships among them from an individual and cross-national perspective. It will also account for the various contextual characteristics that have been found to be relevant for the social capital formation, such as the quality of government, economic development, ethnic composition and welfare state type.

Social Trust, Political Trust and Civic Engagement

Nick Allum; University of Surrey, United Kingdom

Patrick Sturgis; University of Surrey, United Kingdom

Roger Patulny; University of Surrey, United Kingdom

Research on social capital has often focused on interpersonal and political trust and how variation in these constructs across individual citizens and across countries is related to other aspects of social capital. Particular concern has been centred around the relationship between civic engagement and trust (Brehm and Rahn 1997) and also on the relationship between interpersonal trust and political trust (Newton 2001). In this paper we take advantage of structural equation modelling techniques that allow us to integrate both individual and country level information to explore the interrelations between social and political trust and civic engagement. We fit a series of models that begin by estimating the zero-order correlation between these constructs, qua latent variables, and then test for potential mechanisms that might explain the observed associations. Using a range of socio-demographic and personality variables, we find that the oft-cited link between trusting and civic engagement does not stand up well, while the correlation between political trust and interpersonal trust is relatively robust. We conclude by drawing out some conclusions from these results in relation to debates in social capital research.

The Cross-cultural Measurement Equivalence of Generalised Trust in the ESS

Tim Reeskens; University of Leuven, Belgium

Marc Hooghe; University of Leuven, Belgium

In the research literature on social capital, generalized trust features as one of the most important attitudinal elements of social capital and social cohesion in general (Putnam 1993, 2000; Stolle 1999; Uslaner 2002). It is claimed that generalized trust facilitates co-operation with others, and contributes to a general feeling of community and belonging. In comparative research, generalized trust measurements are routinely used to distinguish societies that are high on social capital from those that lack social capital (Kaarianen & Lehtonen 2006; Rothstein & Uslaner 2005; Hooghe, Reeskens & Stolle 2007).

In this paper, we do not wish to get involved in this theoretical discussion. Rather, we want to take a closer look at the validity of generalized trust measurements. If generalized trust is routinely used

in cross-cultural and comparative research, self-evidently we have to be confident that this concept is being measured in a reliable manner across cultures and languages. Despite the fact that the concept of generalized trust has been used so abundantly in international social capital and social cohesion research, the technical aspect of its measurement is seldom questioned. In fact the measurement of generalized trust still largely proceeds using the instruments first developed by Rosenberg (1956). Since 1972 the basic generalized trust questions, as it was introduced in the US General Social Survey has been: "Generally speaking, would you say that most people can be trusted or that you can't be too careful in dealing with people?". Basically, we know that the generalized trust question 'works' in survey research and analysis, and that it is stable over time. But we know less about the content and the precise meaning of the latent variable or attitude it actually refers to.

Especially when looking at generalized trust in a comparative manner, this ambiguity can be problematic. Does the concept have the same meaning for a respondent in the middle of the metropolitan areas of London or Paris, as it has for a respondent in some remote village in the north of Sweden? Basically, this boils down to the question of the cross-cultural measurement equivalence of the investigated variables. With this term, we want to stress the fact that the indicators for measuring the latent concept generalized trust can differ from country to country (Harkness, Van de Vijver & Mohler 2003). Measuring cross-cultural equivalence of latent variables can be implemented by various techniques, such as item analysis, item response theory (IRT), generalizability theory, multi-group structural equation modeling (MGSEM), multidimensional scaling and identity-equivalence methods (Johnson 1998, 25-29). In this article, we will test the cross-national measurement equivalence of the generalized trust items with the MGSEM-technique on the generalized trust items in the European Social Survey (ESS 2002/3; ESS 2004). Previous research has demonstrated that the MGSEM-approach is best suited for testing measurement equivalence across groups and can be preferred above other techniques (Mullen 1995; Cheung & Rensvold 2000; Raju, Laffitte & Byrne 2002; Billiet 2003; Van de Vijver 2003). We hope to conclude with some suggestions on how trust can be measured as reliable as possible in a European, cross-cultural survey.

Testing of a Model of Swift Trust Formation in Temporary Work Groups

Clara L. Popa; Rowan University, United States

Temporary work groups have become a not-so-rare phenomenon in today's organizations (Jarvenpaa & Leidner, 1999). Groups with diversely skilled members who work together on a project for a short period of time and dissolve after the project is completed are a type of resource that organizations employ in order to sustain their competitive advantage. Being mainly a task group, their main goal should be directed toward productivity—having a quality project completed on-time. However, it is the group experience that factors in how organizational members approach groups and their general attitude about working in groups. Trust in group members was found to increase members' satisfaction with the group experience (Boss, 1978; Ward, 1997). In temporary work groups, researchers talk about swift trust as trust that develops fast and holds the group members together. In the present study I propose a definition of swift trust, investigate the mechanisms by which swift trust develops, and propose and test a model of swift trust development in temporary work groups. The study involved 34 temporary work groups who participated in an experiment. The results show that swift trust develops through direct and mediated paths. Affect and proactive attributions of trustworthiness, along with a general predisposition to trust others influence the formation of swift trust. Group communication behaviours—knowledge sharing and suspending judgement—were found to mediate the relationship between the predictors of trust and swift trust. Theoretical and practical implications of the model are discussed.

Analyzing Public Trust in Social Institutions: the Pooled Ordinal Ratings Approach

Tony Tam; Academia Sinica, Taiwan

This paper develops a model-based approach to quantify and statistically explain public trust in social institutions. The approach has three components: (1) measurement assumptions, (2) transformation of

typical data format, and (3) a statistical framework capable of handling responses drawn from repeated cross-sections and multiple responses (trust ratings for multiple targets) from the same respondent.

Surprisingly, only three conventional assumptions implicit in the descriptive item-by-item analysis of trust data are necessary for the approach to work: pooling data across different trust items, aggregating categories of trust (e.g. non-elected government institutions versus elected government institutions), scaling the ordinal ratings simultaneously for all items, and comparing trends of different categories of trust.

The workhorse of our statistical analysis is a random-effects ordered logistic regression model. This model generalizes the conventional ordered logistic regression model by taking into account the potential statistical complications that may arise from the clustering of multiple responses due to the same respondent. This is a simple but flexible model-based framework for studying trends.

An example of important payoffs from successfully pooling the data and simultaneously scaling the ordinal ratings of diverse items is a coherent framework to model the temporal patterns of levels of trust and compare the patterns for different types of trust. I will illustrate the products from the framework with American and Taiwanese data.

Dimensions of Personal Trust and Their Origins

Hajdeja Iglič; University of Ljubljana, Slovenia

The paper is concerned with the question of the dimensionality of personal (social) trust, the relationship between different dimensions and their origins. On the basis of the analysis of Citizenship, Involvement, Democracy survey data for Slovenia, collected in 2002 we distinguish between generalized, particularistic and familiar trust. These types of personal trust are distinguished with respect to the object of trust (people in general, coworkers, comembers, neighbours, friends and family members). While high level of generalized trust is related to high levels of other trust dimensions the vice versa is not true. We also look at different predictors of trust (optimism, interactional and institutional characteristics) in order to show their differential affect on different trust dimensions.

Determinants of Trust in Low Income, Minority Neighborhoods in Texas

Luisa Franzini; University of Texas, United States

The evidence suggests that trust tends to be lower in low income and minority communities. Given the importance of trust to the functioning of society and institutions and to population health, it is particularly important to investigate the determinants of trust in disadvantaged communities. In this paper we investigate the predictors of trust in low-income and minority neighborhoods (defined as census block groups) in Texas. Detailed survey data on 3171 respondents provided information on individual characteristics including sociodemographic and sociocultural personal characteristics (frequency of association with people of other races/ethnicities, social support, perceived racism, perceived personal opportunity, and religiosity). Neighborhood heterogeneities and socioeconomic status, computed from census data, were supplemented by community social characteristics (collective efficacy and public disorder) obtained from survey data. We found that linguistic heterogeneities, but not racial/ethnic heterogeneities, influence general interpersonal trust in disadvantaged neighborhoods. More detailed analyses of generalized and localized interpersonal trust indicate that people tend to trust more those they personally know than those who belong to the same racial/ethnic group. Furthermore, while the results of this study support that people tend to trust more those who are like themselves, the identification of who is more alike is not based exclusively on racial/ethnic identity but is a complex process that varies across races/ethnicities and is based also on linguistic and socioeconomic similarities. Interventions at the neighborhood level should focus on encouraging social interaction in racially/ethnically and linguistically diverse communities in order to promote trust. Through increased social integration among residents, diverse communities can build trust and reach higher levels of cooperation and collaboration that they will contribute to positive economic and health outcomes.

Social Capital and Political Cohesion: How Exceptional is America?

Eric Harrison; City University, United Kingdom

Marc Morjé Howard; Georgetown University, United States

Since the 1990s it has become a commonplace that there has been a decline in levels of 'social capital', as defined in various ways but made most popular in Putnam's work on the US. Similar work on social and political trust has been undertaken in the UK and in Europe. Papers by Newton and Montero (2006) and by Norris (2006) have shown the merits of undertaking comparisons among European countries, but what has been lacking has been data of sufficient equivalence to support true comparisons between America and Europe. This paper combines data from the European Social Survey and the US Citizenship, Involvement and Democracy (CID) survey, each of which fielded a series of items designed to measure the various components of social capital as set out by Putnam. The main purpose of the analysis is to examine how far the evidence supports the thesis of American exceptionalism, or whether the US fits into some broader typology that structures the European countries. The analysis looks at four outcomes - social trust, institutional trust, citizenship attachment and political activity - in relation to the US and eight European countries selected to represent potential ideal-types in terms of social trajectory. The initial results suggest two things: firstly they confirm that many of the constituent elements of 'social capital' are tapping really quite distinct sets of attitudes; secondly they suggest that comparisons between the US and Europe are more nuanced than the thesis of American exceptionalism allows. Indeed some measures place the US mid way along a scale from Scandinavian social democracy to Central European post-communism.

Community Networks Membership, Trust, and Community Engagement: the Case of Local Community Mailing Lists

Ilan Talmud; University of Haifa, Israel

Gustavo S. Mesch; University of Haifa, Israel

This paper inquires into two local community mailing lists operating in Israel. The study analyses the mailing list as a form of 'community networks', (Kavanaugh et al, 2005), exploring their potential effect on community (online and off line) community social capital, as drawn from residents' level of involvement in local activities and organizations. Until recently, Internet research has largely overlooked existing sociological debate related to the impact of technological and societal change on 'community networks'.

In this study we analyse first two waves of a panel study of two community mailing lists, operating in two Israeli localities. Using a survey conducted among 443 residents of these localities, (participants and non-participants of the 'Community Networks' systems), we attempt to explore how trust of local community and off line and online social capital effects residents political efficacy and community involvement.

Findings show that the degree of community residents' civic engagement activities is highly associated with individual's social capital and with Internet linkage. Additionally, we demonstrated the selective nature of community electronic list membership. List members have more pronounced stock of social capital, especially trust in their local organizations, relations and community environment.

Unlike the stipulation put forward by both the utopian and the dys-utopian schools, the Internet does not have an automatic or deterministic effect on community civic engagement. In a "hybrid community", using both offline and online channels to communicate, network infrastructure may enhance existing relational patterns and civic engagement, but neither to replace them nor to abolish their pronounced effect. Internet use is, therefore, embedded with existing social relations. Virtual network may increase the size, strength, and density of the local contacts, but it cannot entirely replace face to face relations.

3.5 European Values Study

Session coordinated by:

Wolfgang Jagodzinski; Central Archive, Germany

Jacques Hagenaars; University of Tilburg, Netherlands

Ruud Luijkx; University of Tilburg, Netherlands

Presentations

1. *Explaining Differences in Economic Cultures: a Multilevel Analysis of EVS Data* – M. Hölscher (United Kingdom)
2. *Heterogeneity in Solidarity Attitudes in Europe. Insights from a Multiple-group Latent-class Factor Approach* – M. Kankarash (Switzerland), G. Moors (Netherlands)
3. *Clustering Countries on the Values Map* – J. Hagenaars (Netherlands), R. Luijkx (Netherlands)
4. *Measuring Religiosity in Intercultural Surveys* – J. Hagenaars (Netherlands), W. Jagodzinski (Germany)
5. *Comparing Comparisons: the Values Studies and Beyond - Cross-national Multilevel Analyses with Different Measurements and Data Sets Testing the Inter-study Stability of Determinants of Attitudes Towards Homosexuality in Europe* – T. Beckers (Germany)
6. *A Cosmopolitan Europe? Who Are the Cosmopolitans and What Orientations Distinguish Them from Others in European Societies?* – F. Pichler (United Kingdom)
7. *Gender Profile on Main Value Orientations in European Societies: the Effect of Social Change* – M. Silvestre Cabrera (Spain), J. Laka Mugarza (Spain), E. Bartolomé Peral (Spain)
8. *Political Support. Explanations in Comparative Perspective in Europe* – E. Bartolomé Peral (Spain)
9. *Civic Participation of Immigrants: Culture Transmission or Assimilation?* – M. Aleksynska (Italy)

Presentations in the session "European Values Study "

Explaining Differences in Economic Cultures: a Multilevel Analysis of EVS Data

Michael Hölscher; University of Oxford, United Kingdom

The European integration process is often discussed in economical terms. Only in the last years, especially in combination with the Turkish application for membership, it became aware that Europe has a cultural component as well. In a comparative project, financed by the German VolkswagenStiftung, we analysed therefore cultural homogeneities and heterogeneities between the countries and people in the enlarged EU (Gerhards 2005; Gerhards and Hölscher 2003). The main data source was the European Values Study from 1999/2000, combining data of a wide variety of subjects and carried out in practically all actual member states of the EU and in Turkey. The results showed that in most value spheres (Religiosity, Family, Politics, Civil Society, parts of Economy) the old EU-member states had the highest affinity to the positions represented by EU organisations. The newer member states showed less conformity, and in Turkey the biggest differences to the official EU positions appeared.

Question for the paper:

The presentation will concentrate on the field of attitudes towards the economy. On the basis of the European Values Study a theoretically derived concept of economic culture was operationalized. Three distinct economic cultures can be identified in the enlarged European Union by using Cluster Analysis and Discriminant Analysis. While we were using a multiple regression to explain the found differences in the values of the people in the overall project, the paper will introduce a more sophisticated analysis of the data in the sphere of economic values by using a multilevel approach (see Hölscher 2006). This allows including explanatory variables on the country level and on the individual level. The results will be interesting both in a substantive and in a methodological perspective. From a substantive perspective the reanalysis will give a much more appropriate explanation of the country differences. From a methodological point of view the occurring differences in the results will give some hints about the limitations and also the robustness of simple multiple regression analyses.

References

- Gerhards, Jürgen (in Zusammenarbeit mit Michael Hölscher). 2005. *Kulturelle Gemeinsamkeiten und Unterschiede in der erweiterten EU. Ein Vergleich zwischen den Mitgliedsländern, den Beitrittskandidaten und der Türkei*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Gerhards, Jürgen, and Michael Hölscher. 2003. "Kulturelle Unterschiede zwischen Mitglieds- und Beitrittsländern der EU. Das Beispiel Familien- und Gleichberechtigungsvorstellungen." *Zeitschrift für Soziologie* 32:206-225.
- Hölscher, Michael. 2006. *Wirtschaftskulturen in der erweiterten EU. Die Einstellungen der Bürgerinnen und Bürger im europäischen Vergleich*. Wiesbaden: VS Verlag.

Heterogeneity in Solidarity Attitudes in Europe. Insights from a Multiple-group Latent-class Factor Approach

Milosh Kankarash; University of Zürich, Switzerland

Guy Moors; University of Tilburg, Netherlands

In cross-cultural and cross-national studies, establishing equivalence in measurement is as important as their reliability and validity. One of most important aspects of equivalence of comparative results is the equivalence in meaning and salience of a construct in all countries studied. This equivalence is in many studies assumed to exist and is not investigated. However, if construct equivalence is not established, biased and misleading conclusions about the nature and significance of the comparison are likely to occur. This paper presents an analysis of construct equivalence for a set of questions concerning solidarity towards different social groups, taken from the 1999/2000 wave of European Value Study. As

a research procedure we used a multigroup latent-class factor approach. Construct inequivalence was confirmed in a model with country as a covariate. Moreover, results demonstrate that models which took into account main sources of construct inequivalence imply substantial changes in estimated effects of countries on latent factors.

Clustering Countries on the Values Map

Jacques Hagenaaers; University of Tilburg, Netherlands

Ruud Lujckx; University of Tilburg, Netherlands

Drawing a values map of the world, locating the countries on this map, is often a key result of cross-national values research. This map is usual two-dimensional based on the two dimensions discovered in the analysis of the separate value items by means of factor analysis or related techniques. The next step then is to ‘cluster’ the countries on this map in ‘homogeneous regions’, where the homogeneity comes from labels such as ‘Confucian countries’, ‘Catholic countries’, ‘English speaking countries’, etc. Given the popularity and the frequent use that is made of this procedure, it may be worthwhile to investigate how this procedure can be improved and made less arbitrary.

This will be done in this paper by focussing on:

- latent class analysis as a more flexible way of discovering the dimensions;
- standard clustering techniques using the ‘loadings’ of the countries on the two dimensions as input;
- probabilistic clustering employing finite mixtures models (latent class models) to cluster the countries on the aggregate or the individual level;
- the association between the cluster ‘scores’ and the scores on external, characterizing variables

Measuring Religiosity in Intercultural Surveys

Jacques Hagenaaers; University of Tilburg, Netherlands

Wolfgang Jagodzinski; Central Archive, Germany

Glock’s seminal distinction between different dimensions of religiosity has not increased the consensus about the measurement of the concept. While some authors conceptualize religiosity as a latent variable with indicators from several dimensions, others have analyzed each dimension separately. Presumably the largest group of sociologists, however, completely ignores underlying variables and latent dimensions and focuses exclusively on the relationships among observed variables. This paper will briefly discuss pros and cons of each approach.

From a methodological point of view, latent variables can and should be introduced if they are theoretically meaningful, simplify the description of the relationship among the observed variables, and correct in the proper way for the unreliability and invalidity of the observed indicators. Further, latent variable models can clarify the discussions about the dimensionality of religiosity by comparing one- and more-dimensional solutions and by investigating whether and how the underlying dimensions are directly linked in a different manner to external variables and/or whether and how particular indicators are directly linked to some external variables (over and above their common relationship with the latent variable(s)).

Here, we will examine by means of latent class and covariance structure analysis

1. whether religiosity can be equivalently measured in different cultures
2. whether or not (or when) religiosity must be treated as a one-dimensional concept and
3. whether the relationship between religious variables on the hand, volunteering and sexual morality on the other hand can be meaningfully described by means of latent variables.

Comparing Comparisons: the Values Studies and Beyond - Cross-national Multilevel Analyses with Different Measurements and Data Sets Testing the Inter-study Stability of Determinants of Attitudes Towards Homosexuality in Europe

Tilo Beckers; University of Köln, Germany

Each cross-national comparative study usually relies on a limited amount of empirical evidence, often just one dataset with probably several measurements. But several items are not always available in one data set. A “one shot” multilevel analysis (using only one item as the dependent variable) has been employed in a study investigating the impact of responsive democracy (Kaufmann), human development (HDI) and gender equality (GDI, GEM) on the attitudes towards homosexuality (Beckers 2006a, 2006b). The results have revealed the strong impact of these country variables, i.e. the contexts (with level-2 variances of 39% and 34% respectively) on attitudes given the world and European data of the 1999-2003 World Values Survey/European Values Study (WVS/EVS). Besides, the results confirm relevant influences of individual predictors of which some have previously been documented in social science research such as age, gender, gender roles, the Inglehart index, education and religion (cf. references in Beckers 2006a).

The WVS/EVS data are limited in three aspects: first, there is a limited measurement of the dependent variable on homosexuality using one item with a ten-point scale, a problem typical for those issues that are not at centre of cross-national research but are still important as contested issues being prototypical for the moral development of modern societies. Second, the dependent variable has skewed distributions in some countries, and third, the item wording is semantically specific (‘is homosexuality justifiable’). Thus, the results ask for validation with different data sets and measurements. This paper focuses on the comparison of several cross-national cross-sectional analyses using data from the WVS/EVS, the International Social Survey Programme (ISSP) 1998 and the European Social Survey (ESS) 2002 in order to test the effects of changing items and distributional patterns on the predicted individual and contextual effects.

Analyses with the WVS/EVS data of 1990 indicate that the causal predictors identified in the initial study are relatively stable over time, including several level-2 and level-1 variables. Furthermore, the different data sets provide a relatively stable set of independent variables but different measurement instruments for the dependent variable, which can be generalized as attitudes towards homosexuality: the ISSP uses the terms sex between two adults of the same sex with an item on a four-point scale while the ESS uses the terms gays and lesbians in a different evaluative context providing a five-point scale. As the data sets also use different level-2 units, i.e. different countries, the computations of the multilevel models are repeated for the same subset of European countries for the three data sets. Although the level-2 variance is partly reduced, it still surmounts many other cross-national analyses. Despite the fuzziness in the measurement instruments provided in the different studies, the country-level indicators are relatively stable and most hypotheses adopted from the initial study are confirmed. The substantial conclusions about the determinants of attitudes towards homosexuality can be generalized as an example of an extended strategy of cross-national comparison in fields of interest providing only limited item choices. The paper finally also raises questions about the standards of comparisons and the opportunities for comparing comparisons in cross-national research given the improved availability of different datasets.

References

- Beckers, Tilo 2006a: Attitudes on Homosexuality in the Context of Human Development. How Gender Equality, Democracy, Religion and Law Shape Moral Tolerance towards Same-Sex Intimacy? A Cross-National Two Level Hierarchical Linear Model. Paper presented at the ‘QMSS-Research Methods Festival’ in Oxford, July 2006.
- Beckers, Tilo 2006b: Das Recht und der homosexuelle Körper in einem weltweiten Vergleich der Zusammenhänge von Rechtsordnung, sozialen Chancen und Einstellungen. Paper presented at

the Plenary Session on 'Körper des Rechts -Körper im Recht' at the 33. Congress of the 'Deutsche Gesellschaft fuer Soziologie' (German Sociological Association) in Kassel, October 2006; to be published in the congress proceedings in 2007.

A Cosmopolitan Europe? Who Are the Cosmopolitans and What Orientations Distinguish Them from Others in European Societies?

Florian Pichler; University of Aberdeen, United Kingdom

This research is concerned with particular developments in individual value and attitude orientations because of cosmopolitanization or 'globalization from within' (Beck 2002) of European societies. Cosmopolitans are thereby defined as people who see themselves and feel like an integral part of the world (Merton 1964 [1949]), or as citizens of the world (Nussbaum 1996). Who are those people? And, more importantly, what distinguishes them from people whose only affiliations are towards people residing within one particular nation-state? With the European Values Study (EVS) it is possible to provide needed insights into these questions. More than any other comparative cross-national survey, the EVS are the data source to capture these developments in Europe over the last 25 years. We compare cosmopolitans to non-cosmopolitans over this period as well as across European countries. Do cosmopolitan identities spread in Europe and in what ways? We assess the cosmopolitans' socio-cultural-economic background using latent class analysis. According to major normative claims of cosmopolitan theory, we expect cosmopolitans to be more open-minded, self-critical and future-oriented, which directly impacts on their values, beliefs and attitudes. Using three-level hierarchical models we further look at decisive differences concerning their values and attitudes towards the environment and the economy, democracy and politics, religion and church, equality and justice, family and gender, migration and finally, future developments over time (level 2) and across countries (level 3) to the extent that the EVS allow this timely comparison.

References

- Beck, U. 2002 'The Cosmopolitan Society and its Enemies' *Theory, Culture & Society* 19(1—2): 17-44.
 Merton, R.K. 1964 [1949] *Social Theory and Social Structure* Revised and enlarged edition. London: The Free Press.
 Nussbaum, M.C. 1996 'Patriotism and Cosmopolitanism' in J. Cohen (ed) *For Love of Country. Debating the Limits of Patriotism* Boston, MA: Beacon Press. pp. 3-17.

Gender Profile on Main Value Orientations in European Societies: the Effect of Social Change

María Silvestre Cabrera; University of Deusto, Spain

Jon Paul Laka Mugarza; University of Deusto, Spain

Eduarne Bartolomé Peral; University of Deusto, Spain

The following proposal tries to describe and explain the variations, in terms of gender, in different European societies, regarding the different aspects of values such as family, religion, politics, work and morality. The cultural background of the different European societies is supposed to play an important role in shaping gender relations, and shaping gender differences in terms of value orientations.

One key aspect is the society or country we observe, and another key element is the moment we are observing. Economic development, modernisation, secularisation, globalisation cause social change and value change and this might play a central role in homogeneity of gender value orientations on the one hand, and also in a slow homogeneity among countries. These hypotheses will be tested.

This proposal will focus on gender and society differences in specific value orientations in order to observe how deep these differences between men and women are and how difference between men and women are varying among countries and over time.

In order to present this communication we will use data of the EVS 1981, 1990 and 1999, and also data from WVS 1995.

Political Support. Explanations in Comparative Perspective in Europe

Eduarne Bartolomé Peral; University of Deusto, Spain

This proposal focuses on political values and orientations of the European citizenry towards their political systems, paying special attention to the concept of political support. This concept, created by David Easton in 1975, tries to measure these orientations of the individuals to their political communities, regime and authorities. These orientations are susceptible to vary depending on the political object (political community, regime, authorities), and also could substantially vary among societies.

The aim of this proposal would be to present, on the one hand, comparisons among European countries regarding the levels of political support, according to the different objects of support. On the other hand, it tries to explain the variations in the effect of the different correlates of political support. Political support may vary depending of the effect of different correlates which also vary in their explanatory capacity depending on the society we are looking at. In order to explain how citizenry articulate political support and how this levels of support to the different political objects are explained by countries, two main explanations will be tested; the explanation of socialization and cultural aspects, and, on the other hand, the short-term explanation of the evaluation of performance, specially economic, as main explanations for the levels and nature of political support. For the analysis I'll be using the data of the European Values Survey 1999.

Civic Participation of Immigrants: Culture Transmission or Assimilation?

Mariya Aleksynska; Università Bocconi, Italy

This paper examines determinants of civic participation of first generation immigrants from fifty-four countries to the European Union using the European Social Survey and World Values Survey datasets. First, the interest is to understand whether culture assimilation with respect to participation takes place. This question is analysed by applying a traditional synthetic cohort methodology, and also by testing whether the levels of immigrants' civic participation depend on the levels of natives' civic participation in the same countries. Second, the paper asks whether the levels of participation of non-migrants in countries of origin matter for participation outcomes of those who migrate, that is, whether culture transmission with respect to participation is happening. In addition, the effect of other country of origin characteristics on immigrants' civic participation is investigated. The issue of immigrants' self-selection is addressed by matching immigrants to otherwise similar natives and compatriots who did not migrate. The study finds limited evidence for the transmission of participation culture across borders, although certain home country characteristics continue influencing participation behaviour of individuals after migration: it is those from industrialized, net immigration, culturally more homogeneous countries who tend to participate more. On the other hand, the culture of current place of residence matters in that by observing higher (lower) participation patterns among natives immigrants tend to participate more (less).

4 Data Analysis

Sessions in Data Analysis

Analysis Strategies for Cross-cultural Research	100
Analysing and Estimating Reciprocal Relationships in Panel Data by Means of Continuous Time Methods	104
Applied Missing Data Strategies in Survey Validation and Use	108
Small Area Estimation: Theory and Applications	112
Categorical Data Arising from Dependent Observations: Multilevel And/or Marginal Models	116
Models of Analysis for Defining and Constructing Subjective Data and Indicators	119
Nonresponse and Measurement Error	124
Testing Structural Equation Models	127

4.1 Analysis Strategies for Cross-cultural Research

Michael Braun; Centre for Survey Research and Methodology (GESIS-ZUMA), Germany

Timothy Johnson; University Of Illinois At Chicago, United States

Presentations

1. *Introduction into the Session* – T. Johnson (United States), M. Braun (Germany)
2. *Comparing the Seemingly Incomparable: the Positional Status Approach to Measuring Achievement and Background* – T. Tam (Taiwan)
3. *Married or Not Married, Religious or Not Religious? Czech and Polish Society According to the ESS Database* – J. Gluza (Poland)
4. *Combining Cross-sectional International Educational Surveys to Assess Achievement Growth: the Case of Pirls and Pisa.* – M. Jakubowski (Italy)
5. *The Political Media Consumer* – M. Fromseier Mortensen (Denmark)
6. *People with Mobility Impairment: a Test of Self-reported Identifiers in Spanish and English* – P. Gallagher (United States), V. Stringfellow (United States)

Presentations in the session "Analysis Strategies for Cross-cultural Research"

Introduction into the Session

Timothy Johnson; University Of Illinois At Chicago, United States

Michael Braun; Centre for Survey Research and Methodology (GESIS-ZUMA), Germany

The introduction gives an overview of the main issues involved in testing for functional equivalence and the main approaches used in the literature. It sets the scene for the contributions to follow. In addition to more advanced techniques, some of the more elementary techniques will be sketched as well. It will be discussed that though the more advanced techniques allow for a higher flexibility and thoroughness in analysis, the elementary techniques are easier to handle for the researcher who is more interested in his or her substantive area than in methodology but they also might allow for a more immediate grasp of the potential pitfalls connected to poor quality of the data.

Comparing the Seemingly Incomparable: the Positional Status Approach to Measuring Achievement and Background

Tony Tam; Academia Sinica, Taiwan

Two profound and pervasive shortcomings have hampered the conventional approach to modeling cognitive ability: scale indeterminacy and incomparability. The paper presents a theoretically grounded approach for circumventing both problems. Particularly useful for comparative research on cognitive inequality, this paper conceptualizes cognitive achievement as the production of positional status and uses the average number of competitors excluded (ACE) as an index of positional status (PSI). The ACE interpretation implies that PSI is a ratio-scale metric with universal comparability. Although the PSI metric is a transformation of percentile scores, PSI has a meaningful zero whereas the meaning a percentile score is no different than a category in an ordinal scale. Ranking order is inherent to the definition of any ordinal scale. It does not contain any meaningful ratio-scale information.

The conceptual and methodological approach is applied to the comparative study of the structure of positional status inequality. The methodology is illustrated by the analysis of two large-scale national surveys of middle school students, the Taiwan Education Panel Survey (TEPS) and the National Education Longitudinal Study (NELS) of the United States. The results show that parental education and income effects on the PSI of cognitive achievement are statistically indistinguishable between the two seemingly very different societies. Moreover, parental education plays a much stronger role than family income in both societies. The remarkable isomorphism of family effects does not only affirm the usefulness of the new framework, but also suggests that the role of parental backgrounds on the positional status of students in cognitive achievement depends much less on the institutional and cultural contexts than on the extent to which cognitive achievement is a form of status competition.

Married or Not Married, Religious or Not Religious? Czech and Polish Society According to the ESS Database

Joanna Gluza; Jagiellonian University, Poland

Despite the fact that Poland and the Czech Republic belong to the same Slavic culture and are neighbours, there are many sociological differences. Dissimilar religion is the fundamental one.

The geographical distribution of believers does not differ within the Polish and Czech regions. A large majority (90%) of Polish society is religious. Moreover, there is hardly any difference between age groups and it seems to be almost unanimous in regions. Czechs represent the complete opposite. Almost 80% of society declares not to be religious, the number of believers increases with age, and differences in regions are more noticeable. The centre of faith is located closer to the Slovakia border moving to the west the religion attitude is changing.

In both countries, religion is strongly related with number of divorces, as well as with number of people living without marriage. It shows that rarely do believers get divorce or stay in relationship without marriage. Existence of that correlation can lead us to the rush conclusion that unbelievers are less likely to be happy in their relationships and that living together without marriage might lead to problems in later relations. Age cohorts analysis reveals other consequence. According to statistics very rarely divorcees decide to get married again. Even those who have found a new live partner tend not to tie the knots again. Presumably the situation is caused by lack of the church approval to marriages after divorce or disbelieves the people luck of believe in the institution of marriage.

Research is based on ESS database and census data of Poland (GUS) and Czech Republic (CSU). Multidimensional statistical analyses were performed with SPSS. It was especially important for the project to show geographical dimension of the problem, that is why various GIS (ArcView) applications have been used to perform spatial analysis and present the data. The comprehensive results of the study are presented on graphs and country maps with region divisions.

Combining Cross-sectional International Educational Surveys to Assess Achievement Growth: the Case of Pirls and Pisa.

Maciej Jakubowski; European University Institute, Italy

This paper proposes a new approach to compare achievement growth in different countries based on international cross-sectional education surveys. The empirical example concerns data from two well known international studies of 4th graders (PIRLS) and 15-year-olds (PISA). The basic idea is that achievement assessed in PIRLS can be used to construct a measure of early achievement for students tested in PISA. Assuming that both studies are representative for the population of students in each country one can compare differences in average achievement in PISA and PIRLS to estimate differences in achievement growth. Such difference-in-differences approach was developed to assess the effects of tracking on student achievement (see Hanushek, Woessmann, 2006). However, this paper shows that such approach is not generally valid if two cross-sectional surveys are measuring achievement of different populations, which is the case for PIRLS and PISA. The paper shows that to assess effectiveness of educational policies cross-sectional data have to be transformed into samples of comparable students. Further, the methodology of matching individual data from two cross-sections to produce dataset with longitudinal information on students is discussed. Several matching methods were examined together with statistical issues of missing data, correcting individual observations and redefining variables to make them comparable across time. The paper shows how the proposed methodology works with empirical data from PIRLS and PISA and alters earlier conclusions about effects of tracking policies in OECD countries.

The Political Media Consumer

Maja Fromseier Mortensen; Danish National Institute of Social Research (SFI), Denmark

This paper is based on data from the second round of ESS surveys. We analyse the consumer of news about politics and current affairs, and the differences in the way these people answer the questions about trust in institutions compared to people less interested in political affairs. Our hypothesis is that people with a fair amount of political media consumption and who in addition to this state, that they are interested in politics, might have other responses to the questions about trust in the different institutions (e.g. the local government, UN or the police). Therefore we have designed "The Political Media Consumer". He is defined as being persons who consume more than 1½ hour per day of political media, and on the same time state that they are very interested or quite interested in politics. The amount of the population classified as political media consumers differs significantly between countries, but generally speaking the political media consumers seems to have more trust in the institutions than the rest of the population. In the paper we present a regression analysis of the relation between the share of the population classified as political media consumers and the average trust in the institutions,

and the results are clear. The higher percentage of the population that can be classified as political media consumers, the higher trust in the institutions can be observed in the country.

People with Mobility Impairment: a Test of Self-reported Identifiers in Spanish and English

Patricia Gallagher; University of Massachusetts Boston, United States

Vickie L. Stringfellow; University of Massachusetts Boston, United States

As part of the ongoing development of the Consumer Assessment of Healthcare Systems & Providers (CAHPS®) survey instruments, the Center for Survey Research is part of a team developing survey items in both English and Spanish for people with mobility impairment (PWMI) to report on the health care they receive.

In 2006, we fielded a dual-language (Spanish and English), dual mode (mail to telephone) survey of members of MassHealth, the Medicaid health insurance program for low-income residents of Massachusetts (n=1,612 completes; AAPOR Response Rate #1= 52%). Candidate PWMI items (see below) were included in all adult instruments. These survey questions included a set of PWMI identifiers (questions #58, 60-63), an open-ended item to allow us to better understand the mobility challenges of people who self identify as a person with impaired mobility (#59), and substantive items about various aspects of care relevant to PWMI (questions #64-67).

Early analysis of the data indicate that overall 31% of respondents report mobility impairment, i.e., either have problems getting around in the home or use mobility equipment, e.g., wheel chair, crutches, etc., or have difficulty walking a quarter mile. Those who elected to respond in Spanish, however, report a significantly higher proportion of PWMI (43%) than those responding in English (29%). We will use administrative claims data (ICD-9 and HCPC codes) available for all sample members for the five years prior to fielding the survey to validate the PWMI identifier items and examine differences by language of the completed instrument.

Goals for analysis are: 1) to examine the cross cultural comparability of the candidate items; 2) to validate the shortened set of PWMI identifiers in Spanish and English; and 3) to examine how the pattern of responses to the items that were included to allow adults with mobility impairment to report on their health care experiences vary by both respondent language and the diagnoses indicated in the claims data.

4.2 Analysing and Estimating Reciprocal Relationships in Panel Data by Means of Continuous Time Methods

Session coordinated by:

Han Oud; Radboud University Nijmegen, Netherlands

Presentations

1. *Spatial Econometric Methods in Continuous Time Modeling of Panel Data with Latent Variables* – H. Oud (Netherlands), H. Folmer (Netherlands), F. van Oort (Netherlands)
2. *Some Conditions for Empirical Identification of Coupled Linear Oscillator Models* – S. Boker (United States)
3. *Continuous Time Modeling of Reciprocal Associations Between Externalizing and Internalizing Problem Behaviors* – M. Delsing (Netherlands), H. Oud (Netherlands)
4. *Three Approaches to Analyse Panel Data: Stochastic Differential Equations, Autoregressive Cross-lagged, and Growth Curve Models* – P. Schmidt (Germany), E. Davidov (Germany), H. Oud (Netherlands)
5. *Non Linear Models for the Feedback Between Gp and Patients* – C. van Dijkum (Netherlands), W. Verheul (Netherlands), J. Bensing (Netherlands)
6. *Assessing the Relationships Between Nationalism, Ethnocentrism, and Individualism in Flanders Using Bergstrom's Approximate Discrete Model* – T. Toharudin (Indonesia), H. Oud (Netherlands), J. Billiet (Belgium)

Presentations in the session "Analysing and Estimating Reciprocal Relationships in Panel Data by Means of Continuous Time Methods"

Spatial Econometric Methods in Continuous Time Modeling of Panel Data with Latent Variables

Han Oud; Radboud University Nijmegen, Netherlands

Henk Folmer; University of Wageningen, Netherlands

Frank van Oort; University of Utrecht, Netherlands

There are compelling arguments for continuous time modelling of panel data, which have discussed by Oud, 2007 (in: Boker & Wenger, Data analytic techniques for dynamical systems). Most social phenomena evolve in continuous time and analysis in discrete time (cross-lagged panel analysis) oversimplifies and often distorts reality. Especially discrete time analysis gets in trouble, when the observation intervals are unequal both within and between studies. Furthermore, a cross-lagged panel analysis offers two sets of coefficients for reciprocal effects between variables, lagged and instantaneous coefficients. In discrete time, there is no way to relate the two different sets with often contradictory results. Perhaps the most compelling reason for analyzing cross-effects in continuous time is that equal effects found in discrete time do not guarantee at all that the underlying continuous time effects are equal. The cross-lagged effects found in discrete time in fact are part of an ongoing process and one should conclude on the basis of the cross-lagged effect functions across time instead of on one single point in time only.

Spatial econometrics has been developed to analyse data, for which the central assumption of independence is not applicable. This is clear especially for spatial data, because less distant sample units in general will correlate higher than more distant units. This kind of dependence is in no way restricted to spatial data, however, but often also present in survey data. We introduce spatial econometric method for solving the unit dependence problem into the latent variable framework and the continuous time modelling of panel data.

Some Conditions for Empirical Identification of Coupled Linear Oscillator Models

Steven Boker; University of Virginia, United States

One model for understanding reciprocal relations in continuous time involves coupled linear oscillators. When two continuously valued processes are considered to be self-regulating such that fluctuations occur about a stable equilibrium and each process may potentially regulate the other process, then coupled linear oscillators may be an appropriate model choice. The parameters from fitting such a model have appealing interpretations that include the natural frequency of oscillations in each process and the strength of coupling between the processes. The coupling can be asymmetric between processes such that process A can have a greater impact on process B than the reverse. In addition, coupling can be expressed in relation to the displacement of a process from equilibrium or in terms of the rate of change in the process. Each of these parameters may have important substantive consequences. The current talk presents results of a simulation that tests the limits of one type of problem that can occur with the linearly coupled oscillator model. When the natural frequency of two oscillators is exactly the same and the oscillators are phase locked, the coupling between the oscillators is unidentified. We explore the range of values of input noise and difference in oscillator frequencies that allow the identification of this model. An example coupled linear oscillator model is presented applied to data from a pilot study of binge eating and ovarian hormones measured daily for 28 days in women with Bulimia Nervosa.

Continuous Time Modeling of Reciprocal Associations Between Externalizing and Internalizing Problem Behaviors

Marc Delsing; University of Utrecht, Netherlands

Han Oud; Radboud University Nijmegen, Netherlands

Over the last decades, several tools have become available for the investigation of reciprocal associations between variables. In the structural equation modelling literature, two methods for the analysis of panel data have received increasing attention: autoregressive models and latent growth curve models (see e.g., Curran & Bollen 2002; Curran & Muthén, 1999; Finkel 1995; Hertzog & Nesselroade, 2003; Jöreskog, 1979; McArdle, 1986, 1988; Meredith & Tisak 1990). Although these methods have important advantages in comparison with cross-sectional designs, they still are problematic in several ways. A problem of both autoregressive cross-lagged models and latent growth curve models, for example, is that they treat time in a discrete rather than in a continuous way. It has been demonstrated that for both autoregressive cross-lagged models (Delsing, Oud, & De Bruyn, 2005; Oud, 2007) and latent growth curve models (Biesanz, Deeb-Sossa, Aubrecht, Bollen, & Curran, 2004), differences in observation intervals lead to different parameter estimates. As a result, outcomes pertaining to different intervals within and between studies are incomparable and often appear to be contradictory (Gollob & Reichardt, 1987; Lorenz, Conger, Simons, & Whitbeck, 1995; Sher, Wood, Wood, & Raskin, 1996). Additional problems of growth curve models are that they do not allow for testing direct causal relations between variables at different time points and that parameter estimates are dependent on the chosen origin relative to which age is scaled (Mehta & West, 2000). The present contribution will demonstrate how a continuous time modeling approach by means of differential equations can resolve the problems of discrete time analysis. Two continuous-time modelling procedures will be described: the approximate discrete model (ADM; Bergstrom, 1966; 1984), and the exact discrete model (EDM; Oud & Jansen, 2000). Both procedures will be applied to the analysis of bidirectional relations between adolescents' internalizing and externalizing problem behaviors. Thus, three theoretical perspectives will be tested: (a) the failure perspective (see e.g., Burke, Loeber, Lahey, & Rathouz, 2005, Capaldi, 1992), which assumes that externalizing problems lead to internalizing problems; (b) the acting out perspective (see e.g., Carlson & Cantwell, 1980, Gold, Mattlin, & Osgood, 1989), which assumes that internalizing problems lead to externalizing problems; and (c) the mutual influence perspective (see Overbeek et al., 2001), which assumes that internalizing and externalizing problems behaviours influence each other reciprocally.

Three Approaches to Analyse Panel Data: Stochastic Differential Equations, Autoregressive Cross-lagged, and Growth Curve Models

Peter Schmidt; University of Giessen, Germany

Eldad Davidov; University of Köln, Germany

Han Oud; Radboud University Nijmegen, Netherlands

To delineate the development over time of authoritarianism and anomia in Germany and their relations, three different approaches for analyzing panel data in structural equation modelling (SEM) are described, applied, and compared: an autoregressive cross-lagged model, a second-order latent growth curve model, and a stochastic differential equation model. Although seldom used, differential equation modeling not only puts real-life processes into a continuous time framework, it also makes the model independent of the arbitrary choice of time points in the data collection. Data was taken from the Group-Focused Enmity Survey, which is a representative sample of the German population. The survey was conducted at three time points between 2002 and 2004. The results of the three models are compared using the same data set from the three-wave panel study. The autoregressive cross-lagged model and the stochastic differential equation model provided support for an effect of authoritarianism on anomia and also slight support for an effect in the opposite direction. The growth curve model provided support for an association between the initial level of anomia and the initial level of authoritarianism. This model also showed a linear increase over time for anomia and no change in authoritarianism. Finally, the merits and disadvantages of the three methods for analyzing panel data are discussed.

Non Linear Models for the Feedback Between Gp and Patients

Cor van Dijkum; University of Utrecht, Netherlands

William Verheul; Netherlands Institute for Health Services Research, Netherlands

Joziën Bensing; Netherlands Institute for Health Services Research, Netherlands

When looking at an ongoing clinical communication between a General Practitioner (GP) and his patients it is evident that different elements do not emerge as a simple function of the participants. Instead the emerging pattern is constantly influenced by the preceding communication. There is a feedback between past and present actions. Moreover there is a feedback between the actions of the participants in the past and the (re) actions in the present.

Simple linear models fail to capture the complexity of those continuous feedback relations. To model the complex recursive causal feedback relations non-linear models have to be used.

With different software it will be demonstrated how such models can generate realistic simulations of real data on Patient-GP interaction. The empirical base of the demonstration and validation is sought in a sample of 102 routine GP hypertension consultations (from a database constituted by the Netherlands Institute for Health Services Research: NIVEL) that were observed, coded with RIAS (a system that uses mutually exclusive categories for every utterance of Patient and GP) and put in data form in SPSS.

Assessing the Relationships Between Nationalism, Ethnocentrism, and Individualism in Flanders Using Bergstrom's Approximate Discrete Model

Toni Toharudin; University of Padjadjaran, Indonesia

Han Oud; Radboud University Nijmegen, Netherlands

Jaak Billiet; University of Leuven, Belgium

In previous research by Billiet (1995 et al., 1995; Billiet et al., 1999), analyses of the Flanders longitudinal 'General Election Study' were done in terms of central theoretical sociological concepts like: ethnocentrism, utilitarian individualism, church involvement, (Billiet, 1995), national identity, subnational awareness, attitude to foreigners (Billiet, Coffé & Maddens, 2005). The present study continues to study the relationships between Ethnocentrism, Individualism, and Nationalism in continuous time. Are these concepts connected by a causal chain or by reciprocal relationships. Practically, it makes a lot of difference, if variables are ordered in a causal chain or influence and enhance each other reciprocally. It is as a result of feedback relationships that systems may show stability and acquire equilibrium states.

The cross-lagged panel design, to be described as a special discrete time case of the state space model, has become an increasingly popular tool to disentangle the over-time influences between variables. However, discrete-time models, such as the discrete-time state-trait model, have important disadvantages. First, the cross-lagged panel design offers two types of effects: cross-lagged effects and instantaneous effects, which often give contradictory results. Second, results from discrete-time analysis can be shown to be highly dependent on the length of the interval between measurements, different time-intervals giving different and contradictory results. The problems are solved by means of the continuous time state space model, using stochastic differential equations. The structural model with the linear constraints, called approximate discrete model (ADM), will be applied. The advantages of the continuous-time state space model in the analysis and consequences of feedback relationships will be illustrated by means of relationships between Ethnocentrism, Individualism and Nationalism in the Flanders election study.

4.3 Applied Missing Data Strategies in Survey Validation and Use

Session coordinated by:

Jason Cole; QualityMetric, United States

Presentations

1. *Missing Data Methods and Applications in Cultural Sociology: Some Examples* – G. Tampubolon (United Kingdom)
2. *Use of Paradata and Simulation to Impute Data on Wage and Salary* – F. Pintaldi (Italy), G. Giuliani (Italy)
3. *Use of Multiple Imputation vs. Last Observation Carried Forward in a Randomized Clinical Trial* – J. Cole (United States), K. Gajria (United States), D. Dubois (Belgium)
4. *A Model for Transferring Variables Between Different Data-sets Based on Imputation of Individual Scores* – B. Todosijevic (Netherlands)

Presentations in the session "Applied Missing Data Strategies in Survey Validation and Use"

Missing Data Methods and Applications in Cultural Sociology: Some Examples

Gindo Tampubolon; University of Manchester, United Kingdom

Burton and Altman (2004) and Horton and Switzer (2005) reviewed the use of methods to deal with ubiquitous missing data in surveys. They found that, despite methodological developments over the last two decades, most applied researchers have not fully considered the impact of missing data on their analyses. If the majority of cancer prognostics studies (Burton and Altman) and clinical studies (Horton and Switzer) are yet to adopt missing data methods, quantitative sociological studies may also need to catch up.

I apply EM-based and multiple imputation methods to deal with missing data prevalent in surveys of cultural consumption or art participation (Tampubolon 2006, Warde et al. 2005). The study begins by delineating a scheme for examining Bourdieu's theory which relates social structure and cultural consumption (the homology thesis). The scheme results in several proposed hypotheses which are tested with a multiple-indicators multiple-causes (MIMIC) model. In the measurement part of the MIMIC model, a conditionally dependent latent class model is used to derive types of cultural consumers; in the structural part, a logistic regression is used to estimate social class effect of being in one of the types. Using survey data sponsored by the Arts Council England (2001 - 2004), it is shown that Bourdieu's homology thesis is supported in the case of music consumption. This is in contrast to other study of the same data which ignores missing data problems.

References

- Burton, A and Altman, DG. 2004. 'Missing covariate data within cancer prognostic studies: A review of current reporting and proposed guidelines.' *British Journal of cancer*. 91:4-8
- Horton, NJ and Switzer, SS. 2005. 'Statistical methods in the Journal.' (Research letter). *New England Journal of Medicine*. 353:1977-1979.
- Tampubolon, G. 2006. 'Distinction in Britain, 2001-2004? Unpacking homology and the aesthetics of the popular class.' Reviewed and resubmitted.
- Warde, A, Tampubolon, G. and Savage, M. 2005. 'Recreation, informal social networks and social capital.' *Journal of Leisure Research*. 37(4):402-425

Use of Paradata and Simulation to Impute Data on Wage and Salary

Federrica Pintaldi; ISTAT, Italy

Gianluca Giuliani; ISTAT, Italy

Wages and salaries by social characteristics (sex, age, occupation and educational level) and also in relation with non-standard forms of paid employment are essential to analyse and understand the labour market. These data will allow analysing how wages and salaries accrue to different categories of employees as a result of their involvement in different kinds of paid employment. We tested the possibility of collecting data on wages and link this information with other relevant information from the LFS. We tested:

- non-response and interview break-offs in different waves;
- the effect of telephone or face to face interviews;
- impact of proxy interviews;
- comparisons among waves between a direct response and a proxy interview;

Concerning the formulation of this variable the tests concerned:

- impact of asking gross or net salary;
- impact of asking exact salary or salary in bands;
- effect of different formulations of the item in different periods or in different waves;
- effect in asking / not asking the respondents to look at their pay-slip.

Analyses also concerned wages for which there is a higher frequency of non-response to this item depending on profession and branch of economic activity. A multivariate analysis technique was used to analyse item non-response and total non-response using other data which is correlated with wages, such as age, educational attainment, profession and economic activity.

The fact of using the data regarding the employees' monthly wage has drawn the attention on the problem of treating the non-responses, which affect almost a fifth of all respondents. We articulated the imputation procedure according to the following phases:

1. Identification and elimination of the outliers;
2. Imputation of the exact value in the case of respondents providing wage in bands;
3. Imputation of missing values.

As regards the identification of the outliers, the purpose was to identify the abnormal values so high or so low that there is a sufficiently high probability they are wrong. Likewise, it is relevant to identify the values that, despite their very low relative frequency, could perturb the estimation of the regression coefficients for imputing the missing values, due to their distance from the distribution centre. Applications by means of the multiivariate techniques will be used.

The imputation phase was carried out through the INPUTE module of the IVEware packet, developed by the Institute for Social Research of the University of Michigan. The module consists of a generalised procedure of multivariate imputation that can treat even relatively difficult data structures under the MAR hypothesis (missing at random). IVEware applies the stochastic regression imputation methods. The results of the methods of regression imputation are strongly influenced by the specification of the regression model and by the set of covariates that we choose to include. In particular, the auxiliary variables must be strongly correlated to wages.

In order to choose the best set of covariates we conducted a simulation using data of respondent to wage and salary. So it was possible to compare imputed data with the answers.

Use of Multiple Imputation vs. Last Observation Carried Forward in a Randomized Clinical Trial

Jason Cole; QualityMetric, United States

Kavita Gajria; QualityMetric, United States

Dominique Dubois; Janssen Pharmaceutica, Belgium

Purpose: Clinical researchers have long asserted that last observation carried forward (LOCF) is a conservative method for analyzing all participants in a clinical trial, including those that drop out. Missing data experts (e.g., Graham, Fairclough, and others) have contested that LOCF is often liberal, rarely justified as appropriately, and almost always inaccurate. For example, presuming a patient who drops out at week 1 will maintain their week 1 outcome value just as a patient who drops out at week 16 will maintain their week 16 value is quite frequently an unrealistic assumption. Multiple imputation (MI) has been posited as an excellent method for treating missingness, including drop out. However, few studies have compared LOCF and MI in randomized clinical trials with anything other than basic statistics.

Methods: Six scales and two summary scores from the Medical Outcomes Study (MOS) Sleep measurement instrument were compared using classic and modern techniques. Data for study came

from a clinical trial comparing once-daily OROS hydromorphone to oxycodone ER for the treatment of patients with osteoarthritis pain. The classic techniques involved using LOCF to replace missing values (including drop out) and change-score analysis to compare group by time differences. Modern techniques involved using the multivariate normal Bayesian MI with mixed model ANOVAs to analyze the group by time differences. Four databases were imputed to obtain efficiency near 95%, and auxiliary variables were used to help ensure the data were Missing At Random (MAR) rather than Missing Not At Random (MNAR).

Results: With classic techniques, two tests were significant at the .05 level, and the mean alpha was .251. With the modern techniques, two tests were significant but the mean alpha was reduced to .181. Additionally, standard error among alpha levels was more than half as small in the modern techniques group (.048) than in the classic techniques group (.113)

Conclusions: Modern missing data techniques provided more accurate and consistent results among patient-reported outcomes in a sleep instrument. With regard to the current clinical trial, the use of modern techniques demonstrated continued improvement of once-daily OROS hydromorphone treated patients over oxycodone ER treated patients.

A Model for Transferring Variables Between Different Data-sets Based on Imputation of Individual Scores

Bojan Todosijevic; University of Twente, Netherlands

It is an often encountered problem that variables of interest are scattered in different data sets. Given the two methodologically similar surveys, a question not asked in one survey could be seen as a special case of missing data problem (Gelman et al., 1998). The paper presents a model for transferring variables between different data-sets applying the procedures for multiple imputation of missing values.

The feasibility of this approach was assessed using two Dutch surveys: Social and Cultural Developments in The Netherlands (SOCON 2000) and the Dutch Election Study (NKO 2002). An imputation model for the left-right ideological self-placement was developed based on the SOCON survey. In the next step, left-right scores were imputed to the respondents from the NKO study. The outcome of the imputation was evaluated, first, by comparing the imputed variables with the left-right scores collected in three waves of the NKO study. Second, the imputed and the original NKO left-right variables are compared in terms of their associations with a broad set of attitudinal variables from the NKO data set.

The results show that one would reach similar conclusions using the original or imputed variable, albeit with the increased risk of making Type II errors.

4.4 Small Area Estimation: Theory and Applications

Session coordinated by:

Tomas Goicoa; Public University of Navarra, Spain

Presentations

1. *Small Area Techniques Applied to Official Statistics* – M. Ugarte Martínez (Spain), T. Goicoa (Spain), A. Militino (Spain)
2. *Estimating Small Area Distribution Functions Using M-quantile and Random Effects Models* – N. Tzavidis (United Kingdom), R. Chambers (Australia)
3. *Out of Sample Estimation for Small Areas* – A. Saei (United Kingdom)
4. *Some Innovations in Prediction for Small Area Statistics' Purposes when One Population Element Can Belong to Many Domains* – T. Zadto (Poland)

Presentations in the session “Small Area Estimation: Theory and Applications”

Small Area Techniques Applied to Official Statistics

M. D. Ugarte Martínez; University of Navarra, Spain

Tomas Goicoa; Public University of Navarra, Spain

Ana Fernández Militino; University of Navarra, Spain

In the past few years, it has been the rule rather than the exception the requirement of local-area statistics by central and local Governments without any change in the sample design originally planned for larger areas. Then, new model-based methods, falling under the headline of ‘Small Area Estimation’, have been developed. The key approach in model-based estimation is to ‘borrow information’ from related areas, past samples or both using implicit or explicit models to connect the small areas through the probability distribution.

It is clear that these techniques are of capital relevance in official statistics, where it is common to provide estimates of a wide range of variables for different administrative divisions. One of the most popular method has been the use of linear mixed models as a basis to obtain empirical best linear unbiased predictors (EBLUP) of population means and totals (see, for example, Rao 2003). Unfortunately, there are some drawbacks of this methodology.

One of them is the difficulty of calculating the MSE of an EBLUP for aggregations of small areas. The EBLUP for the large region, let us say a province, can be obtained as the sum of the EBLUP’s for its small areas, but this is not the case for the MSE. To derive an expression for this MSE is very important because sometimes it is also necessary to obtain estimators in certain administrative divisions, halfway between small areas and provinces (Militino et al., 2007). Another problem has to do with outliers and model specifications. In particular we study the case when there is a single outlying observation in one area, and whether it is appropriate or not to remove it from the sample.

A different, but very important problem in small area estimation is that the estimates obtained for each small area do not add up to the estimate provided by statistical offices at some level of aggregation with appropriate estimators for larger areas. This is known as the “benchmarking” problem, and it is very important because it is related to credibility and consistency of the estimates. We propose to include some restrictions in the estimation process of the model parameters to achieve the desired consistency in the results. We compare this technique with the inclusion of calibration weights used in official statistics by conducting a simulation study.

All the methods are illustrated with real data from the Business Survey of the Basque Country.

References

- Militino, A. F., Ugarte, M. D. and Goicoa, T. (2007). A BLUP Synthetic Versus an EBLUP Estimator: An Empirical Study of a Small Area Estimation Problem. *Journal of Applied Statistics*, 34, 153-165.
- Rao, J. N. K. (2003). *Small Area Estimation*. Wiley Series in Survey Methodology.

Estimating Small Area Distribution Functions Using M-quantile and Random Effects Models

Nikos Tzavidis; University of London, United Kingdom

Ray Chambers; University of Wollongong, Australia

Small area estimation techniques typically rely on regression models that use both covariates and random effects to explain between domain variation. Chambers and Tzavidis (2006) describe a novel approach to small area estimation that is based on modeling quantile-like parameters of the conditional distribution of the target variable given the covariates. This is an outlier robust approach that avoids conventional

Gaussian assumptions and the problems associated with specification of random effects, allowing inter-domain differences to be characterized by the variation of area-specific M-quantile coefficients.

Chambers and Tzavidis (2006) observed that M-quantile estimates of small area means are biased with the magnitude of the bias being related to the presence of outliers in the data. We propose a bias adjustment to the M-quantile small area estimator of the mean that is based on representing this estimator as a functional of the small area distribution function. We consider two approaches. The first approach is based on the use of the empirical distribution function while the second approach is based on the use of the Chambers-Dunstan estimator of the distribution function. Using these two alternative estimators, our method is generalized for estimating other quantiles of the small area distribution function. The effect of this bias adjustment on small area estimation with

random effects models in the presence of model misspecification is examined. Mean squared error estimation for the M-quantile estimates of small area means is also discussed. We conclude by presenting applications with simulated and real data and by providing a summary of current research work.

Keywords: Chambers-Dunstan estimator; Model misspecification; Quantile regression; Robust inference; Robust small area estimation; Weighted least squares;

References

- Breckling, J. and Chambers, R. (1988). M-quantiles. *Biometrika* 75, 761-71.
 Chambers, R. and Tzavidis, N. (2006). M-quantile Models for Small Area Estimation, *Biometrika*, 93, 255-268.
 Koenker, R. and Bassett, G. (1978). Regression quantiles. *Econometrica* 46, 33-50
 Welsh, A.H. and Ronchetti, E. (1998). Bias-calibrated estimation from sample surveys containing outliers, *Journal of the Royal Statistical Society B*, 60, 413-28.

Out of Sample Estimation for Small Areas

Ayoub Saei; Health Protection Agency, United Kingdom

Fay and Herriot (1979) and Battese et al. (1988) are base for indirect small area estimates. In many applications, however, the data structure is more complex than that assumed by Fay and Herriot (1979) and Battese et al. (1988). In particular, there are often no sample observations in some of the small areas (in many applications, the majority of areas) of interest. Clearly, direct estimates cannot be calculated for such out of sample areas. In contrast, model-based estimates for these areas can be computed, but this is usually by making the clearly incorrect assumption that the area random effects for the out of sample areas are zero. This paper extends the empirical best linear unbiased (EBLUP) approach to small area estimation so that estimates for areas in sample as well as those not in sample are calculated in a consistent way. In order to do this, the approach assumes a linear mixed model with spatially correlated area random effects. Estimates of the parameters of the model, including those associated with the random components, are obtained using maximum likelihood and restricted/residual maximum likelihood methods. Mean cross-product error (MCPE) matrix of the small area estimators is obtained. The estimation approach described in the paper is then evaluated by a simulation study, which compares the new approach with other methods of small area estimation for this situation. These results show the new method to be superior.

Key words: EBLUP, out of sample, small area estimation, spatially correlated

Some Innovations in Prediction for Small Area Statistics' Purposes when One Population Element Can Belong to Many Domains

Tomasz Zadło; Karol Adamiecki University of Economics , Poland

The problem of prediction of a domain total value based on the general linear mixed model is studied. In many methods presented in the survey sampling literature a common assumptions are that it is

known for each element of a population to which domain it belongs and that one population element may belong only to one domain. These assumptions are especially important in the situation when a superpopulation model with auxiliary variables is considered. In the paper we propose a method of prediction of the domain total when one element of population may belong to many domains at random with different probabilities. Zero-one random variables are included in general linear mixed model to define to which domains the population elements belong to and additional assumptions on their joint distribution are introduced. Hence, the superpopulation model includes assumptions on the distribution of Y variables and mentioned zero-one variables. Best linear unbiased predictor of domain total and its MSE are derived and their special cases are presented.

Key words: small area estimation, model approach, general linear mixed model

4.5 Categorical Data Arising from Dependent Observations: Multilevel And/or Marginal Models

Session coordinated by:

Marcel Croon; University of Tilburg, Netherlands

Jacques Hagenaars; University of Tilburg, Netherlands

Presentations

1. *Marginal Models for Categorical Data: an Introduction* – M. Croon (Netherlands), J. Hagenaars (Netherlands)
2. *Marginal Models for the Analysis of Panel Data* – T. Rudas (Hungary), W. Bergsma (United Kingdom), R. Németh (Hungary)
3. *Marginal Models for Categorical Data from Multi-actor Surveys* – M. Croon (Netherlands), J. Hagenaars (Netherlands)
4. *Copulas to Deal with Local Item Dependencies* – J. Braeken (Belgium), F. Tuerlinckx (Belgium)

Presentations in the session "Categorical Data Arising from Dependent Observations: Multilevel And/or Marginal Models"

Marginal Models for Categorical Data: an Introduction

Marcel Croon; University of Tilburg, Netherlands

Jacques Hagenaars; University of Tilburg, Netherlands

A lot of social science data arises from research designs that yield dependent observations. Examples are research designs in which several members from the same family are being interviewed or in which the same respondents are interviewed at several points in time. Nowadays, such data are usually analysed by taking the dependencies in the data into account by (re)constructing the data as a multi-level data set and use hierarchical modelling with random coefficients. However, these models are not always optimally suited for all kinds of research questions and usually involve restrictive assumptions about the dependency structure. More often than not, these assumptions remain implicit. Alternative but often overlooked ways of dealing with dependencies in the observations exist, such as transition or regression models, correlated error and autocorrelation models, marginal models (using GEE or MLE), and explicit latent variable models. Nevertheless, the outcomes of these other procedures may yield results, that differ from random coefficient models, especially when dealing with categorical data. And even more important, they may be better suited to answer a researcher's substantive research questions.

In this introductory presentation, marginal models will be further defined and contrasted with a few other approaches to handle dependencies in the data. Further, several older statistical analysis methods will be discussed that can be seen as forms of marginal modelling. Finally, it will be pointed out that marginal models may be relevant in a wider variety of contexts in which dependent observations arise than in longitudinal studies.

Marginal Models for the Analysis of Panel Data

Tamás Rudas; Eötvös Loránd University, Hungary

Wicher Bergsma; London School of Economics, United Kingdom

Renáta Németh; Eötvös Loránd University, Hungary

It is well known that, instead of the separate analyses of the waves of panel data as if they were a series of cross-sectional data, doing analyses on the marginals of the combined data from all waves is more appropriate. When the variables of interest are categorical, regression type analyses are usually performed by fitting log-linear models. Obvious choices of such models involve the same log-linear model for every wave. Such models may be made more relevant by applying the additional assumption of constant log-linear parameter values or a (linear) trend in the parameter values. Further, in addition to the models pertaining to the separate waves, a Markov property of a characteristic across the waves may also be assumed. The talk will discuss the formulation, parametrization and fitting of such models. Illustrative examples from the Hungarian Household Panel will be presented.

Marginal Models for Categorical Data from Multi-actor Surveys

Marcel Croon; University of Tilburg, Netherlands

Jacques Hagenaars; University of Tilburg, Netherlands

In a multi-actor survey several respondents from the same randomly chosen higher level units (families, work teams, schools) are interviewed, inducing in this way dependency among some of the observations. Due to the multilevel structure of the sampling design, standard statistical analyses are no longer appropriate. In stead, one often resorts to random coefficients models which require strong and specific assumptions about the relationships among the variables at the individual level.

In the present paper marginal modelling for categorical data from multi-actor surveys will be discussed. An advantage of this type of models is that it does not require the parameterization of the joint distribution of all variables, but only of those marginal distributions in which one is interested. It will be shown how under this approach maximum likelihood estimates of the parameters of the marginal distributions can be obtained and how the model can be tested. In this approach (Bergsma, 1977) marginal models are formulated as systems of constraints on the cell probabilities and maximum likelihood estimates are obtained via a saddle-point method.

The models and methods discussed in this paper are illustrated on data from the Netherlands Kinship Panel Study (NKPS), a large-scale study into kinship relationships covering a large number of life domains (Dykstra et al., 2004). The data used in this paper come from a module in which a random sample from the Dutch population has been interviewed as well as one randomly parent of each respondent. This particular sampling design creates data that are partly dependent and partly independent. Marginal models are very appropriate for analysing such multi-actor data.

References

- Bergsma, W.P. (1997) Marginal models for categorical data. Tilburg University Press, Tilburg, The Netherlands.
- Dykstra, P.A., Kalmijn, M., Knijn, T.C.M., Komter, A.E., Liefbroer, A.C., & Mulder, C.H.. (2004) Netherlands Kinship Panel Study. A Multi-actor, Multi-method panel study on solidarity in family relationships. Nidi, Den Haag, The Netherlands.

Copulas to Deal with Local Item Dependencies

Johan Braeken; University of Leuven, Belgium

Francis Tuerlinckx; University of Leuven, Belgium

In most measurement situations test items are developed with the intention to provide unique information about the ability or skill being measured. Specific sets of items are, however, often related through a common theme, stimulus material or formulation. Responses on these items will partially show dependency due to their artificial common ground and not only because they relate to the skill or ability intended to be measured. If one wants to get a “pure” assessment of this skill or ability, measurement models should take this effect, often labelled as local item dependency, into account. However, most item response theory models are not robust to local item dependencies, which are in fact violations of the conditional independence assumption underlying most of these models.

Several modelling approaches (e.g., conditioning on other responses, additional random effects) exist that try to incorporate local item dependencies, but they have some drawbacks such as the non-reproducibility of marginal probabilities and resulting interpretation problems. In this contribution, a new class of models making use of copulas to deal with local item dependencies is introduced. These models belong to the bigger class of marginal models in which margins and association structure are modelled separately. It is shown how this approach overcomes some of the problems associated with other local item dependency models.

4.6 Models of Analysis for Defining and Constructing Subjective Data and Indicators

Session coordinated by:

Filomena Maggino; Università degli Studi di Firenze , Italy

Presentations

1. *Subjective Indicators: a Model for Their Construction and Analysis in the Perspective of Data Condensation* – F. Maggino (Italy)
2. *Explaining Satisfaction and Worries by Personality Variables* – B. Rammstedt (Germany)
3. *From Facts to Perceptions: How Do People Recount Their Life?* – C. Laborde (France), G. Vivier (France), E. Lelièvre (France)
4. *The Index of Knowledge Orientation of the Company: an Attempt to Operationalisation* – M. Rószkiewicz (Poland), D. Weziak (Poland)
5. *The Role of Survey Research in Evaluating Regional Development Programmes* – S. Krupnik (Poland), M. Jelonek (Poland), A. Jurczak (Poland)
6. *Advances in Card Sorting Surveys: the Case of Social Issues in the Netherlands* – H. Visser (Netherlands), T. Aalbers (Netherlands), K. van Vringer (Netherlands), D. Verhue (Netherlands)

Presentations in the session “Models of Analysis for Defining and Constructing Subjective Data and Indicators”

Subjective Indicators: a Model for Their Construction and Analysis in the Perspective of Data Condensation

Filomena Maggino; Università degli Studi di Firenze , Italy

The paper shows a model for constructing and analyzing of the subjective indicators. For its characteristics, the model is defined “hierarchical” and is defined by consecutive aspects: a conceptual framework, areas/constructs/domains, variables, items. The application of the model for the definition of the measurement instrument allows to define the questionnaire ‘structured’.

Data analysis allows to condense the data – collected consistently with this model – according to two perspectives:

(A) condensing item information into new synthetic values (indicators) in order to reconstruct the unity of the described concept run through the hierarchical model backwards (→ from items to indicator),

(B) condensing individual values (individual points) in a new meaningful unit (segment) in order to evaluate the observed reality in different perspectives (typological, geographical, etc.) (→ from individual points to segment).

From the operational point of view, the process corresponds to reducing the dimension the data matrix, in terms of number of columns and rows.

In the former perspective, the reducing procedure requires

- the definition of the level of complexity of the condensing process (complexity levels),
- the definition of the criteria of assignment of the weights to each item (weighting criteria),
- the identification of the technique of aggregation of the items in order to define the new indicator (aggregating-over-items techniques).

In the latter perspective, the reducing procedure requires the identification of

- the significant segment (condensing criteria)
- the weight to be assigned to each individual whose indicator/s values have to be condensed in the segment (weighting criteria),
- the technique that allows the aggregation of the individual values of each segment in one or more representative values (aggregating-over-individuals techniques).

That leads to the conclusion that, in our opinion, particular attention has to be paid to the use, to the insertion and to the interpretation of subjective indicators outside their natural context or in different context. The use of one indicator outside his context can lead to a distorted view of the observed reality, for example, using the same indicator whose data are produced in different survey context, in terms of methods for data collection of sample definitions.

Explaining Satisfaction and Worries by Personality Variables

Beatrice Rammstedt; Centre for Survey Research and Methodology (GESIS-ZUMA), Germany

Whether life satisfaction is linked to life circumstances such as unemployment or income, or whether these individual differences can be predicted by stable internal characteristics, or by an interaction of stable characteristics and life events has been discussed for decades (Diener, Suh, Lucas & Smith, 1999; Wilson, 1967). According to Diener’s (1984) assumption of a general tendency to experience life in either a positive or a negative manner, satisfaction can be understood as resulting from a tendency

for positive experiences, while worries might represent the negative manner of experiencing life. This general tendency is supposed to be dependent on the individual's personality. Previous studies have given first support to the notion that indeed the individual's personality might be a potent predictor of life satisfaction (cf. DeNeve & Cooper, 1998). However, it has not yet been investigated in how far personality predicts life worries.

The present study closes this gap by jointly taking personality, satisfaction, and worries into account. Based on a German representative sample of 21,105 respondents, correlations between personality, ten domains of worries and twelve domains of satisfaction were investigated. Worries could be reduced to the three dimensions materialistic worries, post-materialistic worries, and individual worries; the twelve domains of satisfaction were all highly intercorrelated and formed one single satisfaction dimension.

Results clearly showed that satisfaction and individual worries are highly related to personality, while the more external, society-oriented worries are only moderately explained by personality. Regression models for satisfaction and individual worries were highly similar. Not only the degree of variance explained by personality but also the most powerful predictors were nearly identical. It can thus be concluded that personality is a powerful predictor for both individual worries and satisfaction by shaping an underlying general tendency.

From Facts to Perceptions: How Do People Recount Their Life?

Caroline Laborde; Institut National d'Etudes Démographiques, France

Géraldine Vivier; Institut National d'Etudes Démographiques, France

Eva Lelièvre; Institut National d'Etudes Démographiques, France

The study of turning points, critical moments in the life courses of individuals is a long-standing interest in life event history studies, both qualitative and quantitative (Elder, 1977, 1995 ; Ricoeur, 1985 ; Hareven, 1986 ; Leclerc Olive, 1997 ; Thomson et al., 2002, Lelièvre et al., 2006). Depending on the data available, these studies rely on purely factual data or on interpretative statements produced by the respondents. More rarely does the analysis combine both factual information and perceptions which then can be matched.

To go further, we propose to present the factual and subjective quantitative data collected in a French life event history survey conducted in 2001 among 2830 individuals aged 50 to 70 and some first results. Data collected provide both :

- factual events which have occurred in the family, residential and occupational trajectories of the respondents, collected through a Life History Calendar,
- respondents' subjective reconstructions of the main periods of their lives, plus an indicator of happiness during these periods.

This original material allows interesting comparisons between factual events usually collected by researchers and considered significant benchmarks of the life courses (marriages, migrations, parenthood...) and the unique way people recount their lives, identify biographical turning points underlining or not the factual events which have taken place.

The Index of Knowledge Orientation of the Company: an Attempt to Operationalisation

Małgorzata Rószkiewicz; Warsaw School of Economics, Poland

Dorota Weziak; Warsaw School of Economics, Poland

The paper presents the method of creating two indices describing the orientation towards knowledge of the company. The former is the index of the knowledge orientation of a company and consists of three dimensions: (1) acquiring data and information, its updating and dissemination; (2) knowledge generation and dissemination; (3) implementation of knowledge in the decision making process and business activity, as well as sharing knowledge with partners. The latter is the index of the managers'

knowledge orientation. The construction of both indices was based on the model of the knowledge orientation of the company worked out by Mazur, Roszkiewicz, Strzyewska. The model illustrates the relations among managers' knowledge orientation, company's knowledge orientation and organization's performance. Presented article results from the need to verify the model proposed e.g. to measure defined in the model constructs – orientations.

All the computations were based on the data from the randomly chosen sample of 852 companies of 50 – 250 employees. The index of the managers' knowledge orientation as well as each dimension of the index of the knowledge orientation of a company were measured by several (from 14 to 17) 5-point-Likert-type statements. Since both indices were assumed to be latent constructs the choice of the proper method of data treatment was crucial. It was confirmatory factor analysis (CFA). The results obtained were following: (1) the final set of descriptors of all orientations was determined, (2) the coefficients of linear combinations to compute the values of constructs were generated, (3) it turned out that for the companies participating in the research the orientation towards knowledge generation and dissemination is not a significant element of the orientation towards knowledge generally.

The Role of Survey Research in Evaluating Regional Development Programmes

Seweryn Krupnik; Jagiellonian University, Poland

Magdalena Jelonek; Jagiellonian University, Poland

Anna Jurczak; Jagiellonian University, Poland

Regional development is a key issue in European Union due to its role in meeting Lisbon's agenda. It is especially of a great interest in new EU Member States from Central and Eastern Europe, due to significant lack of information within this field and last but not least due to enormous external financing opportunities within current programming period.

However, as a central planning legacy from communist period, regional development programs are exposed to considerable mistrust from potential beneficiaries. One of solutions employed while tackling this problem would be developing a system, which provides regular feedback from all stakeholders in a given region. Beneficiaries' opinion would be treated as a subjective indicator of the undertaken activities. The process of gathering local leaders, investors and inhabitants opinions not only allows a better understanding and stimulates positive attitudes towards regional planning but also promotes higher engagement of those groups within social dialogue.

Provided a single research tool would not be adequate to gather information from such a large and internally diverse population, three versions of the research tool were prepared: for inhabitants, entrepreneurs and representatives of non governmental organisations.

The presentation focuses on developing indicators for Strategy of Regional Development for Małopolska Voivodship. The process had five stages:

1. an analysis of all documents related to the implementation of the regional strategy;
2. desk research of existing indicators of regional growth;
3. in-depth interviews with regional development experts
4. focus groups with representatives of stakeholder's groups: non-governmental organisations, small and medium enterprises and the

inhabitants of the Malopolska region ;

1. a pilot study of the questionnaire drafts
2. preparing the final version of the questionnaires;

The indicators were used both on the regional level and in decomposition within relevant criteria. Being measured in a comparable way, their values may be also evaluated in relation to each other. Thus, indicators values may be compared both within chronological as well as between-group and between-indicators dimensions.

The questionnaire uses also the importance – satisfaction approach toward regional authorities' activities. The presentation presents some challenges researchers faced while applying this technique. Among others, issues involving displaying central tendency and respondents' differentiation.

Advances in Card Sorting Surveys: the Case of Social Issues in the Netherlands

Hans Visser; Netherlands Environmental Assessment Agency (MNP), Netherlands

T.G. Aalbers; Netherlands Environmental Assessment Agency (MNP), Netherlands

Kess van Vringer; Netherlands Environmental Assessment Agency (MNP), Netherlands

D. Verhue; Veldkamp Marketing Research, Netherlands

According to the Social Economic Council of the Netherlands a society becomes more sustainable when consumers and businesses take responsibility for the negative consequences of their activities. In other words, society becomes more sustainable when consumers and businesses contribute to solving important social issues. According to the Dutch government, the social agenda for sustainability must be prominent in major policy-making portfolios. A survey method is presented that establishes the importance that citizens place on social issues. The method applied, a four-stage card sorting survey, makes it possible to prioritize a large number of issues simultaneously.

Three methodological aspects of card sorting surveys will be presented. First, the problem of choosing a proper scoring system will be treated. Then, it will be shown how various survey set-ups can be tested by Monte Carlo simulation (by generating a set of 'pseudo respondents'). Finally, it will be shown that accurate survey-averaged ranking results can be gained without giving respondents all the items/issues present in the survey. To be more specific, if we are interested in the survey-averaged ranks of 64 social issues, we can gain accurate survey-averaged estimates by giving each respondent 20 or even only 2 issues! The only necessary condition is that the issues presented to a respondent are randomly chosen from the total set of 64 issues. Of course, there is a 'cost'. The smaller the amount of items/issues given to a respondent (M), the larger the sample size (N) should be. A relation between M and N will be given in terms of the width of uncertainty ranges for individual issue scores.

The results of the 2006 social-issue survey will shortly be presented (64 issues had to be sorted, 2600 respondents). It is found that Dutch citizens consider the most important issues to be the threat of terrorism, Dutch old-age provisions, combating hunger, human rights and Dutch health care. Many global issues score high: 10 issues out of the top-15 of ranked issues have a global nature, 5 issues have a national nature (this study asked about social issues in general, without placing the focus on the Netherlands or Dutch politics).

4.7 Nonresponse and Measurement Error

Session coordinated by:

Andy Peytchev; RTI International, United States

Presentations

1. *Mode Effects in a Basic Question Approach for the Dutch Labour Force Survey* – F. Cobben (Netherlands)
2. *Relationship Between Measurement Error and Unit Nonresponse in National Probability Surveys: an Approach in the Absence of Validation Data* – A. Peytchev (United States), E. Peytcheva (United States)
3. *Studying Human Populations: a New Curriculum for Social Statistics* – N. Longford (United Kingdom)

Presentations in the session "Nonresponse and Measurement Error"

Mode Effects in a Basic Question Approach for the Dutch Labour Force Survey

Fannie Cobben; Statistics Netherlands (CBS), Netherlands

From July to October 2005, a sample of non-respondents to the Dutch Labour Force Survey (LFS) was re-approached with strongly condensed questionnaires. These questionnaires contained only a small number of basic questions. The main goal of this so-called Basic Question Approach (BQA) is to gain insight into possible differences between respondents and non-respondents with respect to the most important variables of the survey.

Besides the condensed questionnaires, the data collection mode in the BQA was also different from the mode in the regular LFS. The questionnaires were designed for Computer Assisted Telephone Interviewing (CATI) and also Web and paper versions were developed. Non-respondents having a listed land-line telephone number were assigned to CATI, while all other non-respondents received a pre-notification letter with a paper questionnaire attached and a login to the Web questionnaire.

These differences in design complicate a direct comparison of respondents and original non-respondents in the LFS. Possible differences can also be caused by a change of questionnaire and interview mode, i.e. mode effects. Therefore, a fresh control sample was approached with basic questionnaire to determine whether the design of the BQA induced mode effects.

In this paper, we analyse this control sample. We restrict the analysis to the CATI-group because the response to the Web and paper variant was too low. We focus on measurement errors and response bias.

Relationship Between Measurement Error and Unit Nonresponse in National Probability Surveys: an Approach in the Absence of Validation Data

Andy Peytchev; RTI International, United States

Emilia Peytcheva; University of Michigan, United States

Response rates in national probability surveys are falling despite higher cost of data collection from greater field efforts. The inherent threat of lower response rates is the increased potential for nonresponse bias. Some have argued that nonrespondents in a survey would be poor respondents if their cooperation is gained. If true, the higher cost per interview for cases that have already received a lot of effort could be better directed to other areas of the survey process. However, if such respondents do not exhibit more measurement error after controlling for cognitive proxies such as education and age that are can be related also to nonresponse, nonresponse bias should remain as the optimization criterion in decisions during data collection.

Bringing nonresponse and measurement error combines two estimation problems in sample surveys. Nonresponse is a problem of missing survey data, while measurement error requires survey and validation data.

This study uses a model for nonresponse, and proposes a model for measurement error in the absence of auxiliary information. The technique involves the simultaneous estimation of means and variances in purposefully constructed models, and provides respondent-level estimates of measurement error. These estimates provide the potential for studying linkages between survey errors, including the identification of preferable measurement conditions such as particular interviewers inducing less measurement error.

The estimation methods and findings on the relationship between nonresponse and measurement error are demonstrated on two national probability surveys. Particular attention is given to the relationship between nonresponse and means (i.e., bias in point estimates) and between nonresponse and measurement error. Preliminary analysis suggests higher measurement error elicited from respondents rated as less cooperative in a previous survey administration. Implications and needed future developments are discussed.

Studying Human Populations: a New Curriculum for Social Statistics

Nick Longford; SNTL, Leicester, United Kingdom

The presentation, based on a monograph in preparation, describes a new curriculum for social statistics centred around the basic concepts of sampling and measurement. Statistics is defined as making decisions in the presence of uncertainty which arises as a consequence of limited resources at our disposal. By resources, we mean not only finance, but also time, expertise, instruments, goodwill and availability of the respondents, and the like. If these were not limited, the entire population that we study could be enumerated – the values of the key variable(s) established for every one of its members. With limited resources, we have to resort to short-cuts: sampling (observing only a sample from the population) and imperfect measurement (recording the values of a manifest instead of the latent variable). This motivates two central topics: sampling methods and the study of measurement processes.

Statistical models can be introduced as descriptions of infinite populations, which are usually considered as a matter of convenience. To promote integrity, an admission is made upfront that linearity of the models is adopted as a matter of analytical convenience. Sampling methods and model-based estimation represent two distinct perspectives: in sampling, the values of the variables on the population are regarded as fixed, and the act of inclusion into the sample is the sole source of uncertainty; in model-based methods, the values of the outcome variable(s) are regarded as random. The Bayesian paradigm is introduced to complement these two perspectives; in Bayesian analysis, all observed quantities are regarded as fixed and all quantities with unknown values as random.

Non-response or, more generally, failure to collect the data as per plan, is a ubiquitous problem in all large-scale studies of human populations, and methods for dealing with it are indispensable for a statistical analyst. The curriculum follows the innovative research on the EM algorithm and multiple imputation by Don Rubin and his ex-students, and introduces some less standard applications of their missing-data principle. One of them is a general method for dealing with imperfect measurement: the values of the latent variable are treated as the missing information.

With limited resources in mind, the necessity for study design, in all its aspects (sampling, questionnaire, measurement, coding, collation of auxiliary information, and the like) is emphasised. Experimental design is introduced as the setting in which the assignment of the values of some or all of the variables is under the designer's control, and can be randomised. Without it, making causal inferences is very difficult and requires some unverifiable assumptions. In observational studies, to which we resort when experimental design cannot be exercised, the focus rests on the assignment process. The definition of potential outcomes is introduced, connected to sampling methods, and treated as another example of incompleteness.

Advanced or specialised topics include clinical trials, multilevel analysis, generalized linear models and meta-analysis. In general, hypothesis testing is de-emphasised because it is not conducive to making good decisions.

4.8 Testing Structural Equation Models

Session coordinated by:

Albert Satorra; University Pompeu Fabra, Spain

Presentations

1. *Structural Equation Modeling (SEM) for Satisfaction and Dissatisfaction Ratings; Multiple Group Invariance Analysis Across Scales with Different Response Format* – M. Mazaheri (Iran), P. Theuns (Belgium)
2. *The Small Sample Performance of Estimators of the Standard Errors of Structural Equation Models* – J. Lyhagen (Sweden), K. Kraus (Sweden)
3. *Structural Equation Modeling of Intentions to Consume Genetically Modified Food and Pharmaceuticals in Switzerland* – T. Haller (Switzerland), M. Sonneveld (Switzerland), A. Tutkun (Switzerland), B. Lehmann (Switzerland)

Presentations in the session "Testing Structural Equation Models"

Structural Equation Modeling (SEM) for Satisfaction and Dissatisfaction Ratings; Multiple Group Invariance Analysis Across Scales with Different Response Format

Mehrdad Mazaheri; University of Sistan & Baluchestan, Iran

Peter Theuns; Vrije Universiteit Brussel, Belgium

Our focus in current study was, using structural equation modeling, on the development and evaluation of three hypothesized models, LDR, OSWL, and ODWL to predict life domains rating, overall life satisfaction, and overall life dissatisfaction, respectively. A sample of 1310 volunteering students, randomly assigned to six conditions, rated their overall life (dis)satisfaction and their (dis)satisfaction with six different domains of life. Each condition used one of six response formats, differing in (1) orientation (horizontal versus vertical), and (2) anchoring (0 to 10, -5 to +5, and Not Numbered). The results of the confirmatory factor analysis (CFA) support for the conclusion that a six-factor model of LDR can represent participants' response to the Model to Predict Six Life Domain Ratings (LDR) based on the satisfaction and dissatisfaction items. However, as our findings indicate, the kind of response format used for satisfaction and dissatisfaction ratings can affect the factor loadings using confirmatory factor analysis (CFA).

Our results of evaluating the hypothesized models of OSWL and ODWL indicate that the proposed models of OSWL, and ODWL fit the data well, and are able to predict overall satisfaction with life and overall dissatisfaction with life, respectively. Moreover, among six domains of life, modelled as the latent variables, Psychological Well-Being was found to be the strongest predictor of OSWL and ODWL.

The Small Sample Performance of Estimators of the Standard Errors of Structural Equation Models

Johan Lyhagen; University of Uppsala, Sweden

Katrin Kraus; University of Uppsala, Sweden

The small sample performance of estimators of the standard errors of structural equation models When data are analyzed by structural equation modeling, besides parameter estimates and goodness-of-fit statistics, the obtained standard errors are of main interest. With their help it is concluded if parameter estimates are statistically significant and they provide a measure for the stability of the solutions.

However, there has been extensive research on both estimation methods for parameter estimates and on the performance of a multitude of fit statistics, whereas little is done to investigate the properties of different estimates of standard errors. Because several estimators of standard errors are supposed to work equally and thus, are interchangeable for sufficiently large sample sizes, it is of special interest to investigate how they perform for the smaller sample sizes that often occur in empirical research.

We compare four asymptotically equivalent estimators of the variance of the parameters of structural equation models. A simulation study is carried out to evaluate the estimators concerning (i) different distributional assumptions (normal, skew-normal, t, and skew-t) regarding the latent variables and the disturbance terms, (ii) sample size (between 50 and 3200), and (iii) the complexity of the analyzed model. Relative bias, variation and the coverage of confidence intervals are used for the assessment of the estimators. The results show that there are substantial differences in the performance of the four asymptotically equal estimates for standard errors in structural equation models when small sample sizes were analyzed. Most diversity was found for t-distributed, i.e. heavy tailed data whereas skewness has a minor influence on the difference between the quality of the estimators for the standard errors.

Structural Equation Modeling of Intentions to Consume Genetically Modified Food and Pharmaceuticals in Switzerland

Therese Haller; Swiss Federal Institute of Technology Zurich (ETH) , Switzerland

Martijn Sonneveld; Swiss Federal Institute of Technology Zurich (ETH) , Switzerland

Aysel Tutkun; Swiss Federal Institute of Technology Zurich (ETH) , Switzerland

Bernard Lehmann; Swiss Federal Institute of Technology Zurich (ETH) , Switzerland

In surveys, Swiss population expresses a negative attitude towards gene technology in food production. A national vote in favor of a five-year moratorium on production of genetically modified (gm) crops and livestock in 2005 confirms this position. In contrast, biotechnology for medical purpose seems to be valued less critically.

An online survey among students and staff of ETH Zurich, which was conducted in December 2006, reveals the same pattern. The data of this survey (N=2282) served as a base for modeling the influence of attitudes, perceived behavioral control (pbc) and subjective norm on the intentions to consume gm food or gm pharmaceuticals, respectively, as it is proposed by the theory of planned behavior (ToPB, Ajzen 1985). Thereby, a structural equation modeling approach including cross-validation was applied. These basic models for gm food and gm pharmaceuticals were extended with beliefs about the expected benefits and risks of the technology and the consumption of gm products. As expected from a theoretical perspective, these additional explanatory variables have a direct effect on attitude and consequently take an indirect influence on intentions.

In both gm food models (basic and extended version) the influence of pbc on intention was not significant on a 5% level. The intention was mainly explained by the attitude, while the influence of subjective norm is significant but much smaller. By extending the model, the explanation of the intention to consume gm food could not be improved. The modeling of gm pharmaceuticals was more promising, showing significant influences of pbc and subjective norm in the basic model and an improved explanation of the intention by extending the model. Our results suggest, that in the case of hypothetical intentions (supply in gm food in Switzerland can be neglected) the ToPB adds modest insights to the explanation of these intentions because there is little difference between attitude and intention. However, it is useful for more practical intentions (e.g. to consume gm pharmaceuticals), where the addition of risk and benefit beliefs as explanatory variables improves the model.

5 Sampling and Nonresponse

Sessions in Sampling and Nonresponse

Miscellaneous	132
Reducing, Detecting, and Adjusting for Nonresponse Bias	139
How to Cope with Decreasing Response Rates	144
Survey Participation and Non-response Bias in Surveys and Access Panels	146
Selection Bias in Panel Research	149
Probability-based Recruitment for Online Panels	152

5.1 Miscellaneous

Lluís Coromina Soler; ESADE, Universitat Ramon Llull, Spain

Presentations

1. *Measuring Information Accessibility and Predicting Response Effects: the Relative Validity of Differently Transformed Response Latencies* – V. Stocké (Germany)
2. *The Definition of an Optimal Strategy to Set the CATI Call Scheduling Parameters for Households or Individuals Surveys* – S. Macchia (Italy), M. Murgia (Italy)
3. *European Versus Muslim Attitudes Towards the USA* – T. Bechtel (Germany)
4. *Data Collection with Rare Populations: the Example of the German Local Exchange Systems* – S. Wagner (Germany)
5. *Sampling and Measurement Error in Business Surveys Targeting Formal and Informal Businesses* – Z. Ismail (South Africa), Z. Kimmie (South Africa)
6. *A New Model to Classify Femicide: the Italian Case* – D. Iezzi (Italy)
7. *The Concept of Human Rights in the Socio-cultural Context of the Middle East: a Structural Equation Model* – E. Kritski (Canada)
8. *Challenges in Communicating Data Quality for Composite Indices* – F. Cartwright (Canada)

Presentations in the session "Miscellaneous"

Measuring Information Accessibility and Predicting Response Effects: the Relative Validity of Differently Transformed Response Latencies

Volker Stocké; University of Mannheim, Germany

Survey respondents' answers are frequently influenced by seemingly irrelevant features of question wording, the presentation of response options and different aspects of the response context. One explanation for these response effects is the insufficient accessibility of the requested information in respondents' memory. Using different indicators for information accessibility, researchers tested whether respondents' susceptibility to response effects decreases when the cognitive accessibility of the requested information increases. Some studies utilized response latencies, which can be regarded as an innovative accessibility measure in survey research. However, the results about the predictive power of this indicator for response effects are rather mixed. One possible reason for these inconsistent results is the very heterogeneous way latency data is treated before being utilized in data analysis. In order to reduce the characteristic skewness and the impact of outliers, response latencies are often transformed using different statistical functions. Furthermore, different ways of standardizing response latencies were suggested to control for measurement error caused by individually stable differences in the base-line speed and variability. Whether these different treatments affect the response latencies' validity for predicting response effects has not been studied systematically yet.

In our study, we compare the predictive power of raw response latencies for response effects with those of three differently transformed and three differently standardized versions of response times. The analyzed types of transformations are the logarithm, the square root and the reciprocal. Difference-, ratio- and z-scores are included into the study, representing different ways of standardization. The validation criterion is how well the different measures predict the effect of using either high- or low-frequency response scales on the respondents' reports about their daily television consumption. Data from a local random probability sample shows in a first step that raw response latencies are a significant predictor for whether different presentations of response options affect answers: Faster responses are found to be significantly less affected. Second, the different transformations are found to have neither a positive nor a negative effect on the response latencies' predictive power. This is, however, the case for the standardized versions. Here, z-scores, computed using the mean and standard deviation of the individual respondents' response times in the complete interview, proved to be the best predictor, which nearly doubles the raw response latencies' ability to predict the analyzed type of response effects.

Key words: Data Transformation, Information Accessibility, Response Effects, Response Latencies.

The Definition of an Optimal Strategy to Set the CATI Call Scheduling Parameters for Households or Individuals Surveys

Stefania Macchia; ISTAT, Italy

Manuela Murgia; ISTAT, Italy

The growing number of households and individuals CATI surveys in Istat gave us the opportunity to try to determine standards for setting the parameters of the CATI Call Scheduling System. Till now, the in-house development of the electronic questionnaire together with the design of the survey paper questionnaire have led to the definition of standards about the entire structure of the questionnaire (checks, skipping rules and layout) and of its questions (wording, answer categories, etc). Less attention has been paid to the definition of standards for the call management system which has instead a strong impact on the quality of data (response and non response rates) and on the survey budget. This job aims at defining call scheduling standards independent from the target population (households or individuals) and from the fieldwork time.

The analysis is based on three CATI surveys, 'University-to-work transition survey and perspective - 2004', 'Sample Births Survey - 2004' and 'Violence against women - 2006', whose questionnaire has been developed using the Blaise software.

Target populations as well as fieldwork time were different among these surveys:

- the first one was conducted among university graduates and the fieldwork time was 7:30-9:30 pm;
- the second survey was based on a sample of mothers of babies born in 2003-2004 and fieldwork was from 5:30 to 21:30 pm;
- the last survey was run among female household members aged 16 to 70 and the fieldwork was

from 9:00 am to 9:00 pm.

The first step of the analysis was to determine the optimal maximum number of contacts for each unit after which assigning a definitive contact result. This parameter is set a priori in each survey according to its own characteristics and needs. The optimal sought value should allow to reach sample units keeping the response rate at the expected level and avoiding inefficiencies represented by too many calls per unit that lead to negative contact results. We analysed the relative percentage differences among interviews and refusals distributions per number of contacts and we found the optimal maximum value to vary around 6-7 contacts. After this point refusals are significantly much higher than interviews thus lowering the survey productivity. For the 'Violence survey' this value ranges around 3-4 contacts probably because of the extremely delicate subject which led people to deny their cooperation as soon as the interview starts.

The second step of the study was to try to determine the best contact strategy to find out how to distribute the phone calls during the survey period. Results show that if the fieldwork is short (3-5 hours per day) different contacts per unit should be tried in different days in similar time slices instead of in different hours of a single day. If the fieldwork is large (more than 5 hours) more contacts can be made on a single day but with a time difference of at least 2 hours between each other; upper limits for appointments during the same day should also be fixed (max. 3-4 appointments per day).

European Versus Muslim Attitudes Towards the USA

Timothy G. Bechtel; Advis, Germany

For almost a century, and through two world wars, the United States has played a major role in world affairs. After World War II Europeans were grateful for the Marshall Plan which rebuilt their ravaged nations. Muslims now, however, feel their culture and rights infringed by the United States, who has attacked Iraq without provocation and is seen to support Israel in driving fellow believers from Palestine. Is the US motivated by a desire to spread democracy or by greed, appetite for oil, and the needs of empire? What answers do Europeans have, themselves "western brethren" of Americans and earlier missionaries and colonizers?

The present study compares European and Muslim attitudes toward the United States and its motivations. The data are taken from the Pew Global Attitudes Project, which is a nine-nation survey carried out in March 2004 by the Pew Research Center for the People and the Press. The UK, France and Germany are used to represent European attitudes toward the US. Jordan, Morocco, Pakistan and Turkey provide identically measured Muslim attitudes. This dependent variable is a composite of attitude toward George W. Bush and attitude toward the United States. Attitude toward Americans is a control variable in both the European and Muslim regressions.

First, it is observed that both Europeans and Muslims have an unfavourable attitude toward the United States. The Muslim attitude is, however, more extreme. Second, for both Muslims and Europeans this attitude is driven by the opinion that the United States does not consider their interests in making international decisions. Third, opposition to the US war on terror more strongly affects Muslim than European unfavorability toward the United States. Surprisingly, however, Muslim attitude is less driven by perceived US disingenuousness; namely, by perceptions that the US overreacts to terror, is

untrustworthy, and is not really interested in promoting democracy. Europe's greater sensitivity to these factors may stem from its own history of orientalism.

Finally, an important methodological result (reported to ESRA in Barcelona) is confirmed in the present study; specifically, ordinary regression of cardinal attitude scales can replace logistic regression in large national surveys. This is good news for survey researchers who have been coding attitudes numerically for decades.

Data Collection with Rare Populations: the Example of the German Local Exchange Systems

Simone Wagner; University of Konstanz, Germany

Though telephone surveys have become a commonplace in survey research almost no findings concerning its applicability for rare populations exist. Nevertheless social research is extremely interested in studying special or rare populations. However, some specifics of these populations are great challenges for sampling, field access and the choice of survey mode. The total population of these groups is often extremely small, not clearly defined and documented incompletely. Due to their sensitive status cooperation of the target group is not guaranteed. Drawing back on data of a project concerned with "local exchange networks" this presentation points out typical methodological problems investigating special populations, demonstrates techniques for gaining field access and highlights the advantages of the telephone survey for exploring this population. The presentation finishes with some recommendations for special or rare populations in general.

Sampling and Measurement Error in Business Surveys Targeting Formal and Informal Businesses

Zenobia Ismail; Community Agency for Social Enquiry (CASE), South Africa

Zaid Kimmie; Community Agency for Social Enquiry (CASE), South Africa

Conducting business surveys in South Africa and other developing countries is complicated by the absence of accurate sample frames from which to draw probability samples. Among the few accurate and complete sample frames is that of the Johannesburg Securities Exchange, which is limited to public companies and excludes private companies, other formal business entities not operating as companies and the myriad of informal enterprises. However, small, medium and micro-enterprises comprise the majority of business entities in South Africa and are estimated to contribute 37% of the country's GDP (Bradford 2007) and most business surveys, particularly those that address development and economic growth, would have to include these smaller, unregistered enterprises. A number of business surveys undertaken in South Africa have resorted to the use of non-probability samples for the selection of unlisted entities. These selection processes include the use of telephone directories (Muhanna 2007), area sampling based on enterprise density statistics (Bradford 2007) and use of available lists of businesses which are not representative (Jinabhai 2004). In addition to the sampling challenges, surveys which include informal sector enterprises must contend with measurement error difficulties such as scalar equivalence and translation equivalence which can compromise the reliability and validity of cross-cultural comparative research (Sin et al 1999).

This paper will examine the sampling complexities and scalar equivalence of a large customer satisfaction survey of 3 000 enterprises encompassing both formal and informal business entities in the Gauteng province. We will describe and compare the sampling methodologies and their implementation in the formal and informal sectors. The paper will also assess the scalar equivalence of the formal and informal sub-samples, particularly with respect to income and satisfaction rating scales. Reliability analysis and an assessment of content validity are used to compare the use of the scales. The paper proposes alternatives for the design of comparative surveys targeting formal and informal enterprises.

References

- Bradford, W.D. 2007. Distinguishing Economically from Legally Formed Firms: Targeting Business Support to Entrepreneurs in South Africa's Townships. *Journal of Small Business Management*. Vol 45 (1). P 94.
- Jinabhai, M. 2004. Empirical Findings on the Impact of Affirmative Action on Training and Development of Black Managers for Corporate Organisations in South Africa. *Public Personnel Management*. Vol 33 (1). P 121.
- Muhanna, E. 2007. Conceptual Analysis of the Determinants of Entrepreneurship: A South African Perspective. *Problems and Prospects in Management*. Vol 5 (1). P 95.
- Sin, L.Y.M, Cheung, G.W.H and Lee, R. 1999. Methodology in Cross-Cultural Consumer Research: A Review and Critical Assessment. *Journal of International Consumer Marketing*. Vol 11 (4). P 75.

A New Model to Classify Femicide: the Italian Case

D.F. Iezzi; University of Rome "La Sapienza", Italy

Femicide, homicide involving female victims, accounts for approximately one third of all homicides in Italy (1). Despite its prevalence, femicide has received considerably less theoretical attention than the more common male-on-male homicide as criminal offence. The current study uses the Economic and Social Research Centre (E.U.R.E.S.) DB that has collected data on Italian homicide since 2000. The EURES DB presents textual information (description of mechanics of killing), qualitative (type of femicide "in family" or "out of family", relationship between victim and murderer, profession of victim, region of murder, geographical area, size of city, citizenship of victim) and quantitative (year of murder and age of victim) data. The most important information on the mechanism of femicide is textual, collected from newspaper articles. We analysed the 1,125 femicides which occurred between 2000 and 2005 in Italy. The victims are 80% Italian and 20% foreigners. Intimate femicide (764 cases, 67.9%) involved female victims who were killed by their intimate partner in a relationship characterised by jealousy, possessiveness and violence (2).

The aim of this paper is to propose a method to classify the mechanisms of killing and compare the femicides in and out family. It is therefore essential to examine the contexts in which a woman was killed based on the relational distance between victim and offender. In the pre-processing, we selected and tagged (3) the key-topics of textual descriptions of mechanisms of femicide (MF) into a three-way dissimilarity matrix (4, 5). After, we proposed a new method to analyse textual information with qualitative and quantitative data.

The mechanics of femicide in and out family are very similar, e.g. the most used weapons are pistols or guns, followed by knives and the killer hits the woman repeatedly while the victim is already dead. In family, the killer generally strikes the victim on her head or strangles her. Out of the family, he hits her on the head, chest and pubes.

References

- (1) EURES (2006). *L'omicidio volontario in Italia*. In *Rapporto Eures-Ansa 2006*, ed. EURES.
- (2) KILPATRICK D. G. (2004). What Is Violence Against Women? Defining and Measuring the Problem, in *Journal of Interpersonal Violence*, 1209-1234.
- (3) BOLASCO S. (2005). *Statistica testuale e text mining: alcuni paradigmi applicativi*, in *Quaderni di Statistica*, Vol 7.
- (4) BORG I., GROENEN P. (2005). *Modern Multidimensional Scaling, Theory and Applications*. New York, Springer-Verlag.
- (5) LANDAU S. F., HATTIS ROLEF S. (1998). Intimate femicide in Israel: temporal, social, and motivational patterns, *European Journal on Criminal Policy and Research* 6: 75-90.

The Concept of Human Rights in the Socio-cultural Context of the Middle East: a Structural Equation Model

Eugene Kritski; GlobeScan Incorporated, Canada

The presentation will be based on the empirical data collected by GlobeScan Incorporated on behalf of Amnesty International in 2006 among youth in urban areas of Egypt, Jordan, Saudi Arabia and United Arab Emirates. The objective of this study was to investigate how the concept of human rights fits into the value system of Arab youth. The model reveals interrelations between cultural values, perceptions of human rights, and perceptions of Amnesty International as an international human rights advocacy organization and includes the following latent constructs: Repression and violence, Minority rights, Freedom, Traditionalism, Equality, Terror, Trustworthiness of Amnesty International, Cultural relevance of Amnesty's mission, and Impact of Amnesty's activities on national interests.

The model uncovers the archetypal paradigm that defines a perceptual continuum regarding human rights and human rights advocacy groups. This paradigm, stemming from the 'shame culture,' sees traditional values and cultural identity as being challenged by foreign ideas, ideas which tend to be championed by ethnic and religious minorities. The confrontation between traditional values and the individualistic values of the West destabilizes social life and results in violence and terror. The people (the majority) are mainly satisfied with the status quo and believe that tough measures are acceptable for achieving the 'sacred' goal of protecting order, cultural identity, and national pride.

Challenges in Communicating Data Quality for Composite Indices

Fernando Cartwright; Canadian Council on Learning (CCL), Canada

The statistical challenges inherent in the production of a composite index maybe underestimated or even disguised by the apparent simplicity of their interpretation. Indeed, one of the virtues of a composite index is its ability to simplify complex sets of data to a summary that is accessible to lay audiences (Saltelli 2007, Nardo et al., 2005). A composite index is an aggregate measure of statistical indicators taken from a variety of data sources. These sources may be a single survey sample, multiple samples, or even a combination of multiple surveys with administrative data. While communicating technical uncertainty is common with survey-based statistics, it is often neglected in the field of composite indicators (Rotmans and van Asselt 2001; Walker et al., 2003). Consideration of error may not be a priority for indices that are political or illustrative exercises; however, it becomes of paramount importance when indices serve for decision-making (e.g. resource allocation to weak performing areas).

There are two types of error represented in any index: systematic error (bias) and random error (variance). Although a statistically optimal solution attempts to minimize both sources, there tends to be a trade-off between the two. Some methodological choices increase one while decreasing the other. Generally, variance-minimization is preferred in situations where interpretability of results is a high priority. Since interpretability is the primary justification for production of indices, substantial bias is tolerated in order to facilitate straightforward interpretation. Practical consequences of prioritizing variance-minimization include using fixed-effect models, reducing the dimensionality of the indicator set, and using available (rather than ideal) data sources.

In addition to sampling and measurement variance associated with the basic indicators, there is also uncertainty associated with the statistical method(s) used to determine the optimal aggregation procedure of the indicators. The methods with the greatest potential for providing variance estimates for this variance component are based on resampling techniques such as the jackknife and bootstrap. The primary challenge in implementing these methods is choosing the appropriate unit of analysis. Reporting units from different surveys may not be nested, although they are usually geographically reconcilable. Aggregating and disaggregating different variables using geographic models may be necessary to create a rectangular data matrix from which weight coefficients may be estimated.

Practical considerations of these arguments on measurement error and its treatment will be discussed in the context of the Canadian Composite Learning Index (CLI) (Cartwright et al., 2006) that produces a single score of lifelong learning in Canada. A high CLI score means that a particular city or community possesses learning conditions that support economic and social success.

The paper concludes by presenting four recommendations for communicating data quality for composite indices: (i) employ replication methods to estimate variance error at the highest resolution (i.e.

use the smallest unit of reporting as the unit of analysis); (ii) use a common unit of analysis rather than attempting to model pseudo-nested data structures; (iii) identify all sources of bias including original survey questions, choice of indicators, and methodology; and (iv) use the composite indicator as a gateway to accessing the wider set of indicators in a structured manner rather than as a complete summary. The largest remaining challenge lies in developing a methodology that approximates a covariance matrix for multiple indicators across multiple data sources and different units of analysis.

References

- Cartwright F., Mussio J., Boughton C. (2006) Developing the Composite Learning Index - A Framework, Canadian Council on Learning. <http://www.ccl-cca.ca/>
- Nardo, M., Saisana, M., Saltelli, A., Tarantola, S., Hoffman, A., Giovannini, E. (2005) Handbook on constructing composite indicators: methodology and users guide, 2005, OECD-JRC joint publication, OECD Statistics Working Paper, STD/DOC(2005)3, JT00188147, pp. 108.
- Rotmans, J. van Asselt, M.B. A. (2001) Uncertainty in integrated assessment modelling: a labyrinthic path. *Integrated Assessment 2* (2):43-55.
- Saltelli, A. (2007) Composite indicators between analysis and advocacy, *Social Indicators Research*, 81, 65-77.
- Walker, W.E., Harremoës, P. , Rotmans, J. , van der Sluijs, J. P., van Asselt, M. B. A., Janssen, P., Kreyer, M. P., Kreyer von Krauss, M. P. (2003) Defining uncertainty: a conceptual basis for uncertainty management in model-based decision support. *Integrated Assessment 4* (1):5-17.

5.2 Reducing, Detecting, and Adjusting for Nonresponse Bias

Session coordinated by:

Tom W. Smith; University of Chicago, United States

Presentations

1. *Efficiency and Bias in a Two-phase Field Model for Nonresponse* – C. O’Muircheartaigh (United States), S. Eckman (United States)
2. *Representative Wealth Data for Germany from the German Soep: the Impact of Methodological Decisions Around Imputation and the Choice of the Aggregation Unit* – J. Frick (Germany), M. Grabka (Germany), E. Sierminska (Germany)
3. *How General Are Findings About Assumed Non-response Bias in Cross-nation Surveys?* – J. Billiet (Belgium), G. Loosveldt (Belgium), I. Stoop (Netherlands), K. Beullens (Belgium)
4. *The Multi-level Integrated Database Approach (MIDA) for Improving Response Rates: Adjusting for Non-response Error, and Contextualizing Analysis* – T. Smith (United States)
5. *Weighting and Nonresponse in European Social Survey (round 2)* – V. Vehovar (Slovenia), T. Zupaniè (Slovenia)
6. *How Does Length of Fieldwork Period Influence Non-response Phenomenon and Differences Between Respondents and Non-respondents?* – P. Sztabinski (Poland), F. Sztabinski (Poland), D. Przybysz (Poland)
7. *The Effect of a Refusal Avoidance Training (rat) on Final Disposition Codes* – R. Schnell (Germany), M. Trappmann (Germany)
8. *The Effect of Incentives in Reducing Non-response Bias in Multi-actor Surveys* – L. Castiglioni (Germany)

Presentations in the session “Reducing, Detecting, and Adjusting for Nonresponse Bias”

Efficiency and Bias in a Two-phase Field Model for Nonresponse

Colm O’Muircheartaigh; University of Chicago, United States

Stephanie Eckman; University of Chicago, United States

Both Cochran (1963) and Kish (1965) in their books on sample design recommended the subsampling of nonrespondents for more intensive follow-up as an approach to the problem of nonresponse in surveys. There are two arguments in its favour. Where the relative cost of the more intensive procedure is sufficiently high, the process may produce more efficient estimates (measured by cost per effective sample element). Moreover, if the respondents obtained through the more intensive procedure represent more effectively the pool of nonrespondents, the process will also lead to a reduction in nonresponse bias.

The 2006 General Social Survey (GSS 2006) implemented a subsampling design. Using both field records from GSS 2006 and the responses obtained in the survey, the paper will evaluate the effectiveness of the 2006 design, and develop a model to estimate the parameters of a model to guide such designs in the future.

Representative Wealth Data for Germany from the German Soep: the Impact of Methodological Decisions Around Imputation and the Choice of the Aggregation Unit

Joachim R. Frick; Diw Berlin, Germany

Markus M. Grabka; Diw Berlin, Germany

Eva M. Sierminska; Diw Berlin, Germany

The definition and operationalization of wealth information in population surveys and the corresponding microdata requires a wide range of more or less normative assumptions. However, the decisions made in both the pre- and post-data-collection stage may interfere considerably with the substantive research question. Looking at wealth data from the German SOEP, this paper focuses on the impact of collecting information at the individual rather than household level, and on “imputation and editing” as a means of dealing with measurement error.

First, we assess how the choice of unit of aggregation or unit of analysis affects wealth distribution and inequality analysis. Obviously, when measured in “per capita household” terms, wealth is less unequally distributed than at the individual level. This is the result of significant redistribution within households, and also provides evidence of a significant persisting gender wealth gap.

Secondly, we find multiple imputations to be an effective means of coping with selective non-response. There is a significant impact of imputation on the share of wealth holders (increasing on average by 15%) and also on aggregate wealth (plus 30%). However, with respect to inequality, the results are ambiguous. Looking at the major outcome variable for the whole population—net worth—the Gini coefficient decreases, whereas a topsensitive measure doubles. The non-random selectivity built into the missing process and the consideration of this selectivity in the imputation process clearly contributes to this finding. Obviously, the treatment of measurement errors after data collection, especially with respect to the imputation of missing values, affects cross-national comparability and thus may require some cross-national harmonization of the imputation strategies applied to the various national datasets.

Keywords: Wealth, Item non-response, multiple imputations, SOEP

How General Are Findings About Assumed Non-response Bias in Cross-nation Surveys?

Jaak Billiet; University of Leuven, Belgium

Geert Loosveldt; University of Leuven, Belgium

Ineke Stoop; Social and Cultural Planning office (SCP), Netherlands

Koen Beullens; University of Leuven, Belgium

In the first round of the European Social Survey, non-response bias was studied for five countries that had more than 100 converted refusals (reluctant respondents) in their samples. Two of these, Germany and the Netherlands, had even more than 450 reluctant respondents. Information of the extensive contact forms was combined with substantial information in the questionnaire, assuming that the reluctant respondents are informative concerning non-response bias since they are in between the co-operative respondents and the real refusals. In order to assess the plausibility of this hypothesis, both types of respondents (reluctant and co-operative) were compared on a number of (latent) variables that were expected to be sensitive for unit-non-response: political participation, political interest, trust in politics, social trust, attitudes towards immigrants, and voter turnout. With the samples of Germany and the Netherlands multivariate models were tested in order to study the net-effect of type of respondent on the substantive variables. The hypotheses about non-response bias were rejected in some of the country samples (Billiet, Philippens, Fitzgerald & Stoop, 2005). The data of ESS second round 2 contain more reluctant respondents for more countries, including The Netherlands, and Germany. We are thus able to replicate and refine previous analysis and study how far the relations between type of respondents and substantive (attitudinal) variables are comparable between countries and ESS rounds. The results will be evaluated from the viewpoint of survey quality versus survey costs.

Billiet, Jaak, Michel Philippens, Rory Fitzgerald & Ineke Stoop (2005), Estimation of Response bias in the European Social Survey: using information from reluctant respondents in Round one. Paper presented at the 60th AAPOR Conference held in Miami Beach Florida, May 12-15, 2005

The Multi-level Integrated Database Approach (MIDA) for Improving Response Rates: Adjusting for Non-response Error, and Contextualizing Analysis

Tom W. Smith; University of Chicago, United States

We live in an information age in which considerable data exist in many forms and at many levels (e.g. individuals, households, neighbourhoods, communities, etc.). These data sources can be tapped to aid in the collection, adjustment, and analysis of surveys. As Stoop has observed, by linking sampled cases “to a large number of other registers and administrative records, a large amount of data is available on non-respondents and respondents with which non-response can be analysed” and as Zanutto and Zaslavsky have noted, “Administrative records are a relatively inexpensive source of detailed information, especially as technology increases our ability to manipulate large datasets.”

The multi-level, integrated database approach (MIDA) can help to deal with the problems of both low response rates and the detection and adjustment for non-response bias. In addition, MIDA will also further substantive analysis by providing aggregate-level information for contextual analysis. Thus, MIDA will both advance survey-research methodology as well as enhance substantive research.

The essence of MIDA is to use databases to collect as much information as practical about the target sample at both the case-level and at various aggregate levels during the initial sampling stage. The following description of MIDA will use the example of national US samples of households based on addresses and as such is directly appropriate for postal and in-person samples. However, similar approaches can be applied to other modes and populations (e.g. national, RDD, telephone samples; panel studies; list-based samples; and local surveys).

The first step in MIDA is to extract all relevant, public information at both the case-level and aggregate levels from the sampling frame from which the sample addresses are drawn.

The second step is to augment the sampling frame by linking all cases in the sample to other databases at both case and aggregate levels.

The third step in MIDA is to take information gained from the initial case-level linkages to secure additional information.

The final step is to record, process, clean, and maintain a large amount of paradata for each case.

Weighting and Nonresponse in European Social Survey (round 2)

Vasja Vehovar; University of Ljubljana, Slovenia

Tina Zupanič; University of Ljubljana, Slovenia

The data for 24 countries from the second round of European Social Survey were weighted according to Census data for age (three categories), gender and education (three categories). Post-stratification or raking method was applied. Target variables were selected as 45 most typical ESS variables. The weighted and un-weighted estimates (the ESS design weights were always included) were compared - the difference has been attributed to the non-response error.

After weighting, the median relative change of the estimates was calculated and was found to be roughly around 1.4% of their values, but some variables (particularly those related to media, politics, religion) experienced much higher shifts, the highest for newspaper reading (5%). One fifth of estimates (out of $24 \times 45 = 1080$) had a relative bias above 2% and 6% of the estimates have it above 5%;

Usually, the change in weighted estimate was due to the impact of education: in the majority of countries the weighting increased the share of the less educated persons, so the estimates moved towards characteristics of this segment. However, there were also some countries where the share of more educated persons increased after weighting and, correspondingly, this was accompanied with the opposite shift in the estimates. The results were consistent also with the corresponding calculations performed for the first round of the ESS.

Compared to standard errors, the biases achieved $t=1.2$ (median) of the value of the corresponding standard errors. However, in certain topics (politics, media) values were much higher (e.g. "interest in politics" $t=3.6$). Specific variables in some countries even exceeded the value of $t=5$. We should add that the estimates for relative biases are conservative as we assumed SRS sample in calculations of the sampling error.

The results vary considerably across the countries, however, in general it is true that the larger were the non-response rate, the higher were the estimated non-response biases (i.e. differences between the weighted and un-weighted estimates). The results thus warn that a non-response effect does exist in the ESS, particularly for countries with lower response rates. Of course, the weighting only captured one part of non-response bias since only a limited number of control variables (i.e. three variables) was included in the adjustments. In addition, the standard and untested assumption was used that the data were missing at random (MAR).

How Does Length of Fieldwork Period Influence Non-response Phenomenon and Differences Between Respondents and Non-respondents?

Pawel Sztabinski; Polish Academy of Sciences, Poland

Franciszek Sztabinski; Polish Academy of Sciences, Poland

Dariusz Przybysz; Polish Academy of Sciences, Poland

One of possible ways to increase the response rate is to extend the length of fieldwork period. A comparison of data from the Polish ESS pilot study and the main ESS study (Round 2 and Round 3) shows that an extension of the fieldwork period from ca. 2 weeks to over 2 months allows to boost the response rate by 8–10 points in PAPI surveys. This means that some sampled individuals who do not participate in the study when the fieldwork period is short, do take part in it when the fieldwork period is extended.

However, an extension of the fieldwork period also changes the structure of non-response. With a long fieldwork period, the percentage of non-contacts is significantly reduced whereas the decrease of refusals is less visible.

Our presentation shows how the length of the fieldwork period results in differences in responses to socio-demographics and opinion questions between respondents and two categories of non-respondents, i.e. refusers, on the one hand and non-contacted persons on the other.

Data on non-respondents were gathered in a study administered after the ESS pilot study and the ESS main study Round 2. A short mail questionnaire with some ESS questions (socio-demographics

and opinion questions) was mailed to non-respondents, both after the pilot study and after the main study. The mail questionnaire also contained a question about the reason for non-participation in the PAPI study.

The return rate was 52.4% after the pilot study and 36.0% after the main study.

The Effect of a Refusal Avoidance Training (rat) on Final Disposition Codes

Rainer Schnell; University of Konstanz, Germany

Mark Trappmann; University of Konstanz, Germany

Based on Groves/McGonagle (2001) and interviewer manuals of the "Survey Research Center" (Michigan) instructional materials of a German "Refusal Avoidance Training" (RAT) has been developed between 2003-2005. The goal of a RAT is reducing refusal rates by teaching interviewers how to deal with doorstep interactions by maintaining interaction with the respondents and tailoring the arguments to them. The RAT-Training was developed for in person teaching of face-to-face interviewers and as a multimedia self-teaching unit for CATI-interviewers.

The RAT-training has been used for interviewer training in two studies in Germany: For training of face-to-face-interviewers in the German-ESS-Wave 2 (n=5883) and for training of telephone interviewers in the new German "Panel Study Labour Market and Social Security" (n = 12000).

In both studies quasi experimental designs were used to compare final disposition codes between trained and untrained interviewers. The data indicate that RAT-trained interviewers generate significantly fewer refusals, but a part of this nonresponse reduction is compensated by an increase in other nonresponse categories. Nevertheless, a RAT-Training reduced refusal rates in both studies.

The Effect of Incentives in Reducing Non-response Bias in Multi-actor Surveys

Laura Castiglioni; University of Mannheim, Germany

In the pilot study in preparation of the German panel study on families and relationships we implemented a multi-actor design according to which also all partners of our anchors were asked to fill in a written questionnaire, irrespective of the duration and institutionalisation level of the relationship. The questionnaire was generally sent back by mail or, in very few cases, the interviewer could collect it at the end of the interview with the anchor. The response rate was around 55%, which is quite good for a postal survey. Nevertheless, as we feared, longer and more established relationships are over-represented in the partners' survey. Also partners belonging to the youngest cohort (aged 15-17) turned out to be less cooperative.

Thanks to an experimental design, we were able to analyse the effects of offering a 10 Euro conditional incentive to our respondents. The experimental group that was offered an incentive turned out to have a higher response rate, but also the under-representation of not cohabiting partners and less established partnership was reduced. Results hold also when we control for age group.

We suspect that people with stable and established partnerships tend to participate more to a family survey due to their interest in the topic. By offering an incentive we seem to be able to selectively increase the motivation among respondents who do not feel particularly involved in the topic of the survey.

5.3 How to Cope with Decreasing Response Rates

Session coordinated by:

Leo van Doorn; Marketresponse, Netherlands

Presentations

1. *Boosting Response Rates: Are There Optimal Times to Email Respondents?* – S. Bennett-Harper (United States), J. O'Brien (United States), K. Levin (United States), B. Davis (United States), S. Shipp (United States), S. Campbell (United States), R. Sienkiewicz (United States)
2. *Comparison of Some Nonresponse-adjusted Covariance Estimators Under Two-phase Pareto Sampling* – W. Gamrot (Poland)

Presentations in the session "How to Cope with Decreasing Response Rates"

Boosting Response Rates: Are There Optimal Times to Email Respondents?

Sarah Bennett-Harper; Westat, United States

Jennifer O'Brien; Westat, United States

Kerry Levin; Westat, United States

Bryan Davis; Westat, United States

Stephanie Shipp; National Institute of Standards and Technology (NIST), United States

Stephen Campbell; National Institute of Standards and Technology (NIST), United States

Robert Sienkiewicz; National Institute of Standards and Technology (NIST), United States

As response rates to surveys continue to decline (Groves, 2006; Link, Mokdad, Kulp, & Hyon, 2006), researchers have examined a number of factors that might increase respondent cooperation (Singer, Hoewyk, Maher, 2000; Bosnjak & Tuten, 2003). For example, there is a growing body of literature discussing optimal call times for telephone household surveys (Weeks et. al., 1987). With the increasing use of web surveys, it is surprising that there is a paucity of literature about how to increase survey response using paradata, such as the timing of when an email and all reminder notifications are sent to respondents. It may be that when an email notification is sent impacts survey completion rates. It may also be important to send many notifications to prompt survey respondents. In order to address this research question, we alternated the day of the week for sending out email notifications and reminders for a recent web survey conducted for the Advanced Technology Program (ATP) at the National Institute of Standards and Technology (NIST).

For this experiment, individuals were allocated into three groups – email correspondence was sent to Group one on Mondays, Group two on Wednesdays, and Group three on Fridays. Each respondent was sent a survey invitation email and all nonrespondents were sent a maximum of three follow-up reminder emails. Researchers will examine the experimental results to determine: 1) whether a single group generated a higher response rate; and 2) whether the groups differed with respect to the number of reminders required to gain cooperation. Preliminary results suggest that Friday mailings produced higher response rates but required more reminders to achieve this level of cooperation.

Comparison of Some Nonresponse-adjusted Covariance Estimators Under Two-phase Pareto Sampling

Wojciech Gamrot; Karol Adamiecki University of Economics , Poland

Several alternative estimators for the covariance between two population characteristics under double sampling for nonresponse have been considered in the literature. Some approximations of their biases and variances were derived using Taylor linearization. However, the Taylor approximation often needs complicated calculations when complex sampling designs dependent on auxiliary characteristics are used in one or both phases of the survey. Moreover, the quality of these approximations needs verification by itself, especially in the case of relatively small samples where the asymptotic formulas may be highly inaccurate.

In this paper the properties of three alternative estimators of finite population covariance are assessed using simulation experiments. Their biases and variances are compared for the two-phase sampling procedure involving Pareto sampling scheme in both phases.

5.4 Survey Participation and Non-response Bias in Surveys and Access Panels

Session coordinated by:

Peter Lugtig; University of Utrecht, Netherlands

Presentations

1. *Access Panels and Survey Nonresponse: Making It Better or Worse?* – I. Stoop (Netherlands), J. Bethlehem (Netherlands), S. de Bie (Netherlands)
2. *Analysis of Panel Attrition in the German Socio-economic Panel Study Gsoep* – T. Gramlich (Germany)
3. *Nonresponse in Web Based Panel Surveys Applied to Academic Context* – J. Sánchez Carrión (Spain), J. Segovia Guisado (Spain)

Presentations in the session "Survey Participation and Non-response Bias in Surveys and Access Panels"

Access Panels and Survey Nonresponse: Making It Better or Worse?

Ineke Stoop; Social and Cultural Planning office (SCP), Netherlands

Jelke Bethlehem; Statistics Netherlands (CBS), Netherlands

Steven de Bie ; Statistics Netherlands (CBS), Netherlands

Decreasing participation in surveys leads to increasing survey costs or smaller precision because of reduced sample size, and may increase non-response bias in the likely case that noncontacts and refusals differ from respondents. Online panels have been promoted as the solution to this problem as online interviewing is relatively inexpensive, large samples can easily be drawn from panels comprising hundreds of thousands willing respondents and – as internet penetration increases – even specific groups such as the elderly will have internet access and belong to a panel.

Sadly, and despite the rapidly increasing market share of web research, the rising star of online panels is rapidly on the wane. Firstly, response rates of samples drawn from access panels are not always easy to compute as part of the panel may be deadwood – people who have long stopped to participate – and others may be a member under different aliases. This also casts doubt on the size of the panel as advertised. Moreover due to self-selection, selection probabilities are completely unknown, and this makes it impossible to compute unbiased estimates of population characteristics. Secondly, response rates from panel members are decreasing at an astonishing speed. Where in the recent past the high response rates of access panels were the unique selling proposition, response rates of 20% from these access panel members are no exception. In addition, even as internet access becomes almost universal in some western countries, members of access panels are likely to have a number of characteristics – heavy internet user, being online regularly, eager to express one's opinion, experienced with filling in forms, fond of incentives – that may well be related to survey questions and that may cause considerable bias. Finally, there seems to be a high degree of panel overlap. Recent estimates in the US show that an average panel member belongs to 7 other survey panels. Similar evidence from the Netherlands indicates that 80% of the total number of panel questionnaires is completed by 20% of the total number of panel members. Some survey researchers claim that the problems mentioned above can be reduced by applying some kind of weighting adjustment procedure, e.g. using weighting variables measured in a reference survey. We argue this is a too optimistic point of view.

The presentation will outline why access panels do not solve the problems caused by non-response, will discuss access panels in the framework of total survey error and will give examples of when access panels can be useful for academic and governmental research.

Analysis of Panel Attrition in the German Socio-economic Panel Study Gsoep

Tobias Gramlich; University of Konstanz, Germany

Nonresponse is a severe problem in sample surveys, especially in panel surveys. The threat of possible nonresponse bias is complicated by the fact that nonrespondents can differ not only systematically from respondents, but also different types of nonrespondents can differ systematically, splitting the original sample not only in respondents and nonrespondents but in different types of nonresponse categories. Each of the different types of nonresponse can be a source of a possible nonresponse bias and has distinct effects, for example, refusals may bias the mean downward, whereas noncontacts bias the mean away from zero. If nonresponse is not analysed separately, possible effects of the different types can be mixed, making it difficult or impossible to account for possible nonresponse bias accurately.

This paper analyses nonresponse in the German Socio-Economic Panel Study (GSOEP). The GSOEP is the biggest non-governmental scientific panel study of households in Germany. It started 1984 consisting of about 5.900 households and more than 12.000 individuals aged 16 and older within these

households. Since households are not stable and are changing over time, it is individuals who are followed over time.

In the yearly gross data of individuals, the GSOEP documents up to 29 different reasons for nonresponse; some of them lead to final drop out (deceased, moved abroad, explicit refusal), whereas other nonrespondents are contacted the next year again according to the tracking rules. After two consecutive waves with nonresponse, an individual is removed from the panel. Unfortunately, due to changes in fieldwork coding schemes or different coding behavior of the interviewers, the coding scheme for the 29 causes for nonresponse is not consistent over time, yielding 10 distinctive and consistent categories of nonresponse over time.

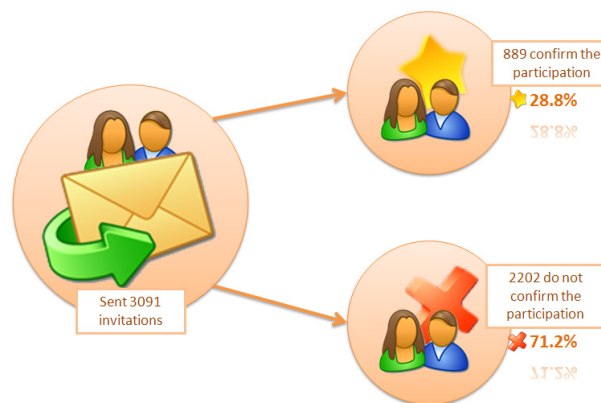
The paper reports the development of the different categories of nonresponse in the GSOEP over time. Furthermore, rational choice theory and a theory of social integration are used in a multinomial logistic regression over time to estimate the probabilities of different nonresponse types.

Nonresponse in Web Based Panel Surveys Applied to Academic Context

Juan Javier Sánchez Carrión; Universidad Complutense de Madrid, Spain

José Manuel Segovia Guisado; Universidad Complutense de Madrid, Spain

In the last two years the Inspección de Servicios of the Universidad Complutense de Madrid (UCM) has been running a web-based panel survey with a probability sample of 889 students. In each wave of the panel students report on the incidences occurred at the academic lectures they attend in a particular day of the week. For this academic year we have a sample of 14 639, being the later our unit of analysis.



In this paper we will report on the nonresponse rate of the panel, on one hand related to those invited to participate (3 091) who declined our invitation (2 202) and on the other to the specific nonresponse rate in the different waves of the panel. We will cover the following aspects:

- Basic methodological characteristics of the research and aim of the survey.
- Nonresponse participation rate in the invitation process and in the different waves of the panel.
- Socio demographic characteristics of non respondents.
- Nonresponse and the recalling process (up to 3 reminders in each wave). Reminders and its incidence in the measurement of basic parameters of the survey.
- Bias as a result of nonresponse
- Implications of the results for online based research in academic context.

5.5 Selection Bias in Panel Research

Session coordinated by:

Peter Lugtig; University of Utrecht, Netherlands

Gerty Lensvelt-Mulders; University of Utrecht, Netherlands

Presentations

1. *A Comparison of Different Methods to Correct for Self-selection Bias in Panel Surveys* – P. Lugtig (Netherlands), G. Lensvelt-Mulders (Netherlands)
2. *Selective Non-response in Constructing an Internet Panel Based on Probability Sampling* – A. Hoogendoorn (Netherlands), J. Daalmans (Netherlands)
3. *Evaluation of the Representativity of an Opt-in Online Access Panel* – G. Loosveldt (Belgium), N. Sonck (Belgium)

Presentations in the session "Selection Bias in Panel Research"

A Comparison of Different Methods to Correct for Self-selection Bias in Panel Surveys

Peter Lugtig; University of Utrecht, Netherlands

Gerty Lensvelt-Mulders; University of Utrecht, Netherlands

Internet based household panels are becoming increasingly popular as an instrument to obtain knowledge of the general population on a wide variety of topics. Since the Internet penetration in the general population is already large and growing (83% in the Netherlands), such panels are often treated as a random sample of the population and the results of panel studies are commonly generalized to the whole population. In this presentation we study the effects of choosing the Internet as only source of data collection on panel quality.

A random sample of 10.000 individuals was drawn from the general population and invited to become a panel member. From the non-responders a second sample of 2000 was drawn and asked to return a short survey containing the same demographic variables and 9 questions from the first panel survey on healthcare reforms.

Results indicate that panel members differ markedly from non-members on almost all demographic variables, and consequently on the means and variances of the knowledge questions.

Different methods are used to correct for these differences. One weighting technique using logistic regression weights and further three forms of propensity matching; simple matching on propensity scores, stratified matching using the propensity caliper and matching on a combination of propensity scores and Mahalanobis distance.

Selective Non-response in Constructing an Internet Panel Based on Probability Sampling

Adriaan W. Hoogendoorn; Vrije Universiteit Amsterdam, Netherlands

Jacco Daalmans; Statistics Netherlands (CBS), Netherlands

Evaluation of the Representativity of an Opt-in Online Access Panel

Geert Loosveldt; University of Leuven, Belgium

Nathalie Sonck; University of Leuven, Belgium

It can be observed that a lot of opinion research is currently carried out through the Internet. Following Couper's typology (2000), one of the eight online survey types is the volunteer panel of Internet users. This opt-in online access panel can be defined as a 'previously recruited poll of respondents who have agreed to take part in surveys' (Stenbjerre & Laugesen, 2005). The commonly known advantages and disadvantages of online surveys over other modes also apply for this type of Internet panel research. In a short fieldwork period and with a very low cost, a large group of people can be accessed, of which the answers can be rapidly collected and analysed because of an extensive automation. However, besides these benefits online research has to cope with some important methodological problems which are mainly associated with coverage (possible bias because a group of people does not have Internet access) and sample selection (no known chance of selection, possible bias from recruiting methods) (Bandilla et al., 2003; Dillman, 2000; Couper 2000). According to Couper (2000) it is most problematic when surveys based on volunteer panels of Internet users claim they are representative of a more general public.

The primary research question of the paper is to what degree data collected by an opt-in online access panel are representative for the general population. To address this question, an empirical comparison is performed between the data obtained by a self-administered online survey with panel members on the one hand and the answers obtained by a face-to-face interview with a sample of the general population

on the other hand. The online panel under investigation is composed of respondents who are partly self-selected (e.g. pop-ups) and partly recruited by non-probability based methods (e.g. client databases), both online and offline. Based on personal background and contact information gathered through a registration procedure, a database of potential respondents is established from which panel members are sampled.

The online panel survey and face-to-face interview are compared based on questions with identical wordings, asked during a similar time period in the same geographic region, namely Flanders in Belgium. Based on findings of previous research (Bandilla et al., 2003) it can be expected that Internet users who substantially differ from the general population in socio-demographic terms and in their interest levels to participate in surveys, they answer differently on various opinion questions. The degree and direction of these differences are considered.

References

- Bandilla, W., Bosnjak, M. & Altdorfer, P. (2003). Survey Administration Effects? A Comparison of Web-Based and Traditional Written Self-Administered Surveys Using the ISSP Environmental Module, *Social Science Computer Review*, 21, 235-243.
- Stenbjørre, M. & Laugesen, J.N. (2005). Conducting Representative Online Research; A Summary of Five Years of Learnings, Paper presented at ESOMAR Worldwide Panel Research Conference in Budapest, April 2005.
- Couper, M.P. (2000). Web Surveys. A Review of Issues and Approaches. *Public Opinion Quarterly*, 64 (4), 464-94.
- Dillman, D.A. (2000). *Mail and Internet Surveys. The tailored Design Method*. New York: John Wiley & Sons, 464p.

5.6 Probability-based Recruitment for Online Panels

Session coordinated by:

Mick Couper; University of Michigan, United States

Presentations

1. *Recruiting a Sample for an Online Panel: Effects of Contact Mode, Incentives and Information Letter* – A. Scherpenzeel (Netherlands)
2. *Differences Between Respondents and Nonrespondents in an Internet Survey Recruited from a Face-to-face Survey* – W. Bandilla (Germany), M. Blohm (Germany), L. Kaczmirek (Germany), W. Neubarth (Germany)
3. *Using an RDD Survey to Recruit Online Panel Respondents* – M. Couper (United States), J. Dominitz (United States)
4. *Who Wants to Become an Online Panel Member? Effects of Respondent Characteristics on the Recruitment of an Online Panel* – C. Vis (Netherlands)
5. *Panel Motivation: Why Panel Members Participate (or Not) in Surveys of an Online Panel.* – M. Marchand (Netherlands)

Presentations in the session "Probability-based Recruitment for Online Panels"

Recruiting a Sample for an Online Panel: Effects of Contact Mode, Incentives and Information Letter

Anette Scherpenzeel; CentERdata, Netherlands

Preceding the recruitment for a new, online household panel a pilot study was carried out, designed to find the optimal recruitment strategy. The factors to be optimised by the experiment were the contact mode, the incentive level, the timing of the incentive, the content of the information letter, and the timing of the panel participation request. Additionally, the experimental design takes the presence of a fixed landline in the households into account, which is a "naturally" varying factor.

Results show that the first choice with respect to incentive payment is a prepaid 10 euros incentive. The four incentive levels tested in the experiment, 0 euros; 10 euros; 20 euros and 50 euros, show a curvilinear effect on response rates. The major response increment is found between 0 and 10 euros; 20 euros and 50 euros incentives do not substantially increase response rates beyond the 10 euros level. All incentives have much stronger effects when included with the announcement letter then when paid afterwards. The second most important factor is the contact mode, which was telephone (CATI) or face-to-face (CAPI). Whereas the contact rate is somewhat higher in CATI, the CAPI interviewers are more successful in obtaining panel participation. The contact rate is substantially lower in the subpopulation of households without a fixed landline, even when controlled for the effect of contact mode.

The announcement letter, which was sent prior to the first contact by the interviewer, varied in the degree to which the text contained special persuasive concepts. The response rates were not affected by the content of the letter. Finally, the nature of the panel study was in some conditions explained to respondents prior to the request to participate in the recruitment-interview, and in other conditions only after the interview has been done. This manipulation did not influence the response rates.

Differences Between Respondents and Nonrespondents in an Internet Survey Recruited from a Face-to-face Survey

Wolfgang Bandilla; Centre for Survey Research and Methodology (GESIS-ZUMA), Germany

Michael Blohm; Centre for Survey Research and Methodology (GESIS-ZUMA), Germany

Lars Kaczmirek; Centre for Survey Research and Methodology (GESIS-ZUMA), Germany

Wolfgang Neubarth; Centre for Survey Research and Methodology (GESIS-ZUMA), Germany

Internet surveys have emerged as a promising data collection methodology. However, not enough is known about differences between people who do and who do not participate in such online surveys, particularly with regard to the general population. In our study aimed at exploring systematic differences between respondents and nonrespondents, we used the German General Social Survey (ALLBUS) as a sampling frame. The ALLBUS was conducted face-to-face in 2006 and achieved a 41% response rate. ALLBUS respondents with Internet access were asked to participate in a Web-based survey. Those willing to participate were then invited by email. Around N=1600 of all ALLBUS respondents (46%) reported having private Internet access. Of those respondents with Internet access, 37% expressed their willingness to participate (N = 588), and were invited by email to complete the Web survey. There were different reminders to improve the response rate in the Web survey: a) an email reminder, b) a postcard prenotifier in combination with a second email reminder, and c) a telephone reminder for those who had not answered after two months. A special group with incorrect email-addresses (N = 70) was invited by letter to do the Web survey. Overall, 64% of those actually invited completed the Web survey. Willingness to participate and actual completion was related to a number of demographic variables. We will report on these differences at the stage of ALLBUS recruiting and the different reminder stages and discuss their implications for recruiting participants for Internet surveys.

Using an RDD Survey to Recruit Online Panel Respondents

Mick Couper; University of Michigan, United States

Jeffrey Dominitz; RAND , United States

Over the past few years, the University of Michigan and RAND have been evaluating the feasibility of Web data collection among older Americans. As part of this effort, we have been using the Survey of Consumer Attitudes (SCA), a monthly RDD telephone survey, to recruit participants for the American Life Panel (ALP), an online Internet panel of persons 40 and older, maintained by RAND. This paper describes these recruitment efforts, and examines demographic and attitudinal correlates of Internet access, willingness to participate in the online panel, and actual participation across several waves of data collection.

Of the almost 7,300 eligible respondents interviewed in the SCA from January 2002 to December 2006, 69% reported using the Internet. About 50% of Internet users expressed willingness to join the ALP, while 42% of non-users agreed. Those who did not use the Internet, along with a subset of Internet users, were assigned to receive telephone surveys. We focus here on the online surveys among Internet users. Internet users who were willing to participate in the ALP were further asked to provide an e-mail address; 82% did so, while 5% reported they had no e-mail address, and 12% declined. A subset of these SCA respondents was sent a mailed invitation to participate in the first online survey for ALP; those who provided an e-mail address were also sent an e-mail invitation. Of those invited, 48% completed the first online survey. We report on the differential coverage and nonresponse errors at each stage of the recruitment process, and also examine retention across several waves of the online panel.

Who Wants to Become an Online Panel Member? Effects of Respondent Characteristics on the Recruitment of an Online Panel

Corrie Vis; CentERdata, Netherlands

Preceding the recruitment for a new, online household panel a pilot study was carried out designed to find the optimal recruitment strategy. A probability sample of addresses was drawn for this study, including households without Internet access and households without a fixed landline. Households without Internet access are provided with Internet and – if necessary – with a computer by the survey institute to be able to participate in the online panel. All households were contacted in a traditional way: first, an announcement letter was sent, explaining the nature of the study; next, respondents were contacted by an interviewer, either by telephone or face-to-face. The interviewer asked them to participate in a 10-minute interview, after which, at the end of the interview, the request to participate in the panel was made. The 10-minute recruitment interview was designed, on the one hand, to give respondents an impression of what the panel interviews would be like, and, on the other hand, to measure the characteristics of respondents who agree and disagree to participate in the panel. Of special interest were characteristics other than the standard demographics, which can often be corrected by weighing, and therefore the interview included indicators of health; social integration; political interest; leisure activities; survey attitudes; loneliness and personality.

Results show, firstly, important effects of age and prior access to Internet. These related effects indicate that for the real recruitment, considerable extra effort has to be put into persuasion of the elderly who are reluctant towards the free Internet access and computer offered by the survey institute. Furthermore, the respondents who were willing to become a panel member were, on average, more interested in the news; made more cinema or museum visits; were more sportive or participating in unions or clubs; more active voters; more self-confident; less lonely; less religious; and more positive towards survey research in general than the respondents who were unwilling to become a panel member.

The rapidly growing subpopulation of respondents without fixed landline can be characterised as of younger age; single person households; working; much less than average involved in news, politics

and societal activities; often non-voters; somewhat more often speaking another language than Dutch at home.

Panel Motivation: Why Panel Members Participate (or Not) in Surveys of an Online Panel.

Miquelle Marchand; CentERdata, Netherlands

It is important to investigate how respondents ‘stay on board’ as panel members, because of the costs involved in replacement of panel members and because high long-term response rates are important for research purposes (representativity and longitudinal analyses). In this paper, we report a study about motivation to participate in a panel and in panel questionnaires. The panel we studied is an Internet based panel, the so-called “CentERpanel”, which consists of about 2000 households in the Netherlands. Every week, the members of those households complete a questionnaire on the Internet, with panel member without Internet access completing their questionnaires with a so-called set-top box.

The main motivation of respondents to participate in the CentERpanel is to help others (researchers), to give their opinion, to learn things, or to help society. Surprisingly, questionnaire characteristics such as length, layout, and difficulty are considered less important than privacy guarantee and scientific relevance. In line with that, the reasons people gave for quitting the panel have little to do with the difficulty or variation in the questionnaires. Most often, respondent quit for personal reasons such as lack of time or they just do not feel like it anymore.

Furthermore, we compared long-term panel members (in the panel for over 437 weeks) with short-term panel members (in the panel less than 179 weeks). The response pattern of these two groups differed on several points: Long-term panel members more often say they always fill out their questionnaires on the same day and time, whereas short-term panel members do not have a fixed moment for this; they say they fill out questionnaires when they happen to think about it and interrupt questionnaires more often. In addition, the short-term panel members more often motivate their participation on monetary grounds (to earn “CentERpoints”).

Summarizing, it appears that motivations to participate are somewhat more intrinsic and habitual for long-term respondents than for short-term respondents.

6 Mode of Data Collection and Data Enhancement

Sessions in Mode of Data Collection and Data Enhancement

Perils and Promises of Web Surveys	158
From Questionnaires to Data File: the Role of the Data Producers and Disseminators . . .	162
The Use of Paradata for Evaluating and Improving Survey Quality	169
Using Contact Forms for Data Quality Assessment of Cross-national Surveys	175
Mobile Phone Usage in Survey Research: Implications for Data Quality	178
Mixing Modes of Data Collection	183

6.1 Perils and Promises of Web Surveys

Katja Lozar Manfreda; University of Ljubljana, Slovenia

Vasja Vehovar; University of Ljubljana, Slovenia

Presentations

1. *Recruitment for Volunteer Surveys and Selection Bias* – K. Tijdens (Netherlands)
2. *Calibration and Propensity Score Weighting in Web Surveys* – E. Fabrizi (Italy), S. Biffignandi (Italy)
3. *Access Panels: a Blessing or a Serious Threat to Political Processes?* – W. Saris (Spain), D. Oberski (Spain)
4. *The Effects of Incentives in Web-based Research: a Review* – A. Göritz (Germany)
5. *Improving Data Quality in Web Surveys with Visual Analogue Scales* – F. Funke (Germany), U. Reips (Switzerland)

Presentations in the session "Perils and Promises of Web Surveys"

Recruitment for Volunteer Surveys and Selection Bias

Kea Tijdens; University of Amsterdam, Netherlands

The WageIndicator is an international, continuous web-survey on work and wages in 19 countries. The web-survey is available at frequently visited WageIndicator websites, which jointly received almost 8 million unique visitors in 2006. The Salary Checker on the web-sites is the major attraction. Worldwide, people show a great desire for information about wages, as the numbers of visitors as well as their emails show. The websites are deeply rooted in the Internet. Many other websites link to WageIndicator websites. They are easily found in search engines. They have a well developed marketing strategy including cooperation with major players on the Internet. See www.wageindicator.org/.

The development of Internet has changed the nature of volunteer surveys. Increasingly volunteer web-surveys are being explored because provided sufficient marketing, the response is overwhelming. Volunteer web-surveys offer unprecedented opportunities in terms of sample sizes, multi-country and multi-lingual coverage, and continuity, all at lower costs than comparable probability surveys. These volunteer web-surveys require web-marketing, or as Faas and Schoen (2006) state 'Putting a Questionnaire on the Web is not Enough'.

This paper discusses the three-step self-selection bias in the volunteer WageIndicator web-survey. First, Internet access causes bias, which may be related to wages, the prime variable in the data. However, the online population in countries with higher Internet penetration tends to be much more diverse and more similar to the larger populations of their countries than countries with lower Internet penetration (Randall, 2007). Second, although the numbers of visitors of the WageIndicator websites are large and growing, it is still a minority of the extremely large and heterogeneous population visiting the Internet. These visits may be related to interest in wages, and therefore maybe high-wage earners could be under-represented. Third, when visiting a WageIndicator website, a visitor has to decide whether or not to complete the questionnaire. This decision may be related to time availability, and therefore visitors with long working hours might be under-represented.

These assumptions have been investigated for the web-survey data for the years 2002-2006. A detailed comparison of socio-demo variables between the web-survey data and the national Labour Force Survey data does not support the first and the third assumption, but it does support the second assumption. The comparison shows that workers in small part-time jobs and workers aged 55 and over are severely underrepresented, but that underrepresentation decreased over the years. As for education, the low skilled labour force is under-represented in the sample. From 2002 to 2004, however, the under-representation reduced.

Calibration and Propensity Score Weighting in Web Surveys

Enrico Fabrizi; University of Bergamo, Italy

Silvia Biffignandi; University of Bergamo, Italy

Web surveys are usually characterized by massive non response and undercoverage. A popular way to deal with these problem in the sampling literature is by means of weighting observations in the estimation process. If the population totals of some auxiliary variables are accurately known post-stratification or, more in general, calibration methods can be used to reduce both non-response and undercoverage bias (Sarndal and Lundstrom, 2005). They are based on the correction of a set of 'basic weights', i.e. the inverse of the inclusion probabilities when sampling is random.

Alternatively an estimated propensity score is sometimes used to create a weight adjustment to account for nonresponse in surveys where some variables are known for both respondents and non-respondents (Little and Rubin, 1987).

When correction of basic weights is substantial the resulting weights may be characterized by high variability and a skewed distribution with extreme values. This may have a bad impact of the mean

square error of estimators. In fact the reduction in the bias may be overcome by an increase in the variance. This is true especially when non response is noninformative, that is the probability of being a respondent is independent of the study variables.

Compromise solutions may be obtained by means of weight trimming or smoothing (Elliot and Little, 2000). In the case of calibration we may also modify the distance function for the same purposes.

In this paper we deal with several problems related to weighting, such as variable selection in building propensity scores and calibration, comparison of different trimming threshold and smoothing methods. We are also interested in evaluating the impact of 'basic' or 'initial' weights on final weights. This problem is relevant for those web surveys in which it is not possible to obtain basic weights as inverse inclusion probabilities.

We illustrate these problems using data from a web survey conducted on a population of university students. More detail we consider a survey on a stratified random sample of students from the University of Bergamo, contacted via e-mail and asked to complete a customer satisfaction questionnaire on the University library services on the internet.

Access Panels: a Blessing or a Serious Threat to Political Processes?

Willem E. Saris; ESADE, Universitat Ramon Llull, Spain

Daniel Oberski; ESADE, Universitat Ramon Llull, Spain

In this paper we discuss the recent paradigm change in survey research. Where survey research so far was based on probability sampling, nowadays many surveys are based on self-selected access panels. This means that the statistical basis for generalization from the sample to the population has gone. Some alternative procedures, like weighting based on propensity scores, have been suggested to correct for possible biases. However, the empirical studies done to compare the results of access panels and probability samples are so far inconclusive. Sometimes large differences are found and it was not possible to correct for the differences, while at other times such differences do not surface.

Information from surveys plays a role in several different political processes such as voting, participation in elections, agenda setting, opinion forming about issues; in general survey play a role in any process which involves a reaction to the perceived general political orientation of the public.

Given that one can not say at this moment whether an access panel provides proper information about characteristics of the population, one runs the risk that political processes are influenced by incorrect information based on access panels.

This leads us to the conclusion for the moment to avoid access panels and rely on probability samples or panels based on a probability sample. Furthermore, we conclude that further research is necessary to study the reasons for participation in panels and possible correction methods for panel bias such as (propensity score) weighting, imputation, likelihood-based estimation, etc.

The Effects of Incentives in Web-based Research: a Review

Anja Göritz ; Friedrich-Alexander Universität Erlangen-Nürnberg, Germany

In view of the importance of Web-based data collection, it is vital to develop methods for promoting the quality of gathered data. By handing out incentives for participating in a study, researchers can influence people's likelihood of taking part in studies and the quality of their responses. Besides incentives, there are other response-enhancing methods such as prenotification, assurances of anonymity, stating a deadline, social-utility- and help-the-researcher-appeals, and reminders. However, meta-analyses routinely show that at least in traditional survey modes incentives have a larger impact than other response-enhancing methods. Therefore, in the relatively new field of online data collection it is reasonable to give the study of incentives priority over the study of other response-enhancing methods. Furthermore, studying the impact of incentives in Web-based studies is necessary because results gained in other survey modes might not apply to the online realm. Because employing incentives might entail both desired and undesired effects, we need comprehensive insights about the effects of incentives, including

trade-offs between different facets of data quality. The paper gives an overview of five years of research on the use of incentives in Web-based studies. The paper reviews a meta-analysis on the topic as well as numerous stand-alone experiments. The paper develops evidence-based guidelines for short-term as well as long-term employment of incentives in online studies to attain the goal of collecting high-quality data in a cost-conscious manner.

Improving Data Quality in Web Surveys with Visual Analogue Scales

Frederik Funke; none, Germany

Ulf-Dietrich Reips; University of Zürich, Switzerland

A Visual Analogue Scale (VAS) is a graphical rating scale, a plain horizontal line with both ends anchored. Respondents indicate the extent of a subjective impression by placing a mark on the line. VAS lead a shadowy existence. This is mainly because of two reasons. First, VAS are used rarely in paper and pencil studies because of the extra work required in measuring the respondents' markings. Second, only little is known about the impact on data quality and the measurement error when using VAS.

With computerized surveys, the first reason for not using VAS does not longer apply. Read out happens automatically, fast and accurate. Using Web surveys the surveying of large samples with VAS is feasible. To explore the impact on data quality and the measurement error of VAS, we conducted four Web experiments, evaluating the quality of data obtained with VAS (also see Reips & Funke, in press).

In the first experiment, we were able to show that data from VAS meet the criteria of an interval scale, e.g. equidistance. Thus, a range of important statistical procedures may safely be applied to data from VAS (e.g. the computation of means, parametric procedures like the t-test, Pearson correlation and linear regression models).

In experiments two and three, we compared different categorical (Likert type) scales to VAS. Our main finding is: Categorical scales do differ systematically from interval level, producing data on an ordinal level only. Therefore, to compare data from VAS with categorical scales, a special transformation is needed that will be described in the presentation.

Experiment four examined retest reliability of VAS and categorical scales. A 40 item personality inventory was repeatedly administered. Result: Although VAS facilitate a far more precise judgment, there was no negative influence on retest reliability. The implications of this series of experiments will be discussed, particularly for using VAS as a way of collecting data on the Web.

6.2 From Questionnaires to Data File: the Role of the Data Producers and Disseminators

Session coordinated by:

Kirstine Kolsrud; Norwegian Social Science Data Services (NSD), Norway

Presentations

1. *The Work and Challenges of the Data Archive* – K. Kolsrud (Norway)
2. *Indicators of Difficulty* – I. Stoop (Netherlands)
3. *Data Quality in the European Social Survey: the Case of Item-nonresponse* – M. Blohm (Germany), A. Koch (Germany)
4. *Effects of Survey Sponsorship and Mode of Administration on Respondents' Answers About Their Racial Attitudes* – V. Stocké (Germany)
5. *Collecting and Providing Proper Documentation for International Surveys: the Example of the ISSP* – M. Quandt (Germany)
6. *Data Editing and Quality Control: Value for Money or Art for Art's Sake? the Polish Experience with the ESS* – Z. Sawiński (Poland)
7. *Clean Data or Cleaned Data? Data Editing Procedures and Experiences of the ESS Data Archive* – K. Kalgraff Skjåk (Norway)
8. *A Method for Appending Annual Surveys* – J. Pasquini (Italy), R. Gini (Italy), R. Berni (Italy), S. Benocci (Italy), S. Forni (Italy), F. Collini (Italy), F. Voller (Italy), E. Buiatti (Italy)
9. *Bazaar Style Data and Metadata in the Age of the Data Cube: a Few Ideas Based on the European Social Survey (ESS)* – H. Orten (Norway)
10. *On the Importance of Meta-data and Documentation of Cross-national Surveys* – P. Mohler (Germany), B. Pennell (United States)

Presentations in the session "From Questionnaires to Data File: the Role of the Data Producers and Disseminators"

The Work and Challenges of the Data Archive

Kirstine Kolsrud; Norwegian Social Science Data Services (NSD), Norway

A traditional view on data archives and data archiving would be an image of a large data repository, a place where researchers, research projects and other data producers deposit their data for eternal preservation, and where researchers could either visit and retrieve the data they sought on site, or they could order data by mail. The work carried out by the archives could be described as a "black box" where the end users would have only a vague idea of what happened with the datasets from they "left" the data collectors to the data files re-appeared available from the archive.

Whereas the reservoir of data is still the foundation of data archives, modern technology in the form of communication technology, the spread of powerful PCs and the development of increasingly user friendly software for advanced statistical analysis, all contribute to a change in the role and the scope of the archive work. The new possibilities provided by the developed technology have also raised additional demands on the archives.

Parallel to the technological advances there has also been a marked development in the spectre of data collected by research projects. Large comparative survey projects like the European Social Survey have increased the focus on quality in all segments of the survey process, resulting in a wide range of data (paradata) from call record data, data on interviewers, sample data, methodological data as well as large amounts of documentation.

The increase in the amount and variety of data adds to the archives' challenges of distributing data and metadata to the users in a user friendly and coherent manner. The distribution of data and meta data directly over the internet has over the last years become the primary means of dissemination for a number of archives (ZA, ICPSR, NSD and others). Whereas the internet greatly improves the accessibility of data and meta data, the same accessibility accentuates the need for comprehensive documentation for secondary analysts who do not have the inside knowledge and collective memory of the primary data producers. The need for comprehensive documentation is also essential for cross national projects where the need for information is many fold the need of the single country studies.

This presentation will use the archive work of the ESS as an example on how NSD have tried to increase the transparency of the archive work, and how we have tried to meet the combined challenges of modern ICT and large scale cross national data collections in our archive and dissemination work.

Indicators of Difficulty

Ineke Stoop; Social and Cultural Planning office (SCP), Netherlands

In analyzing nonresponse, 'difficult' respondents can be used as a proxy for final nonrespondents. These difficult respondents are often subdivided into those that are hard to reach and those that are initially reluctant to cooperate. The former are assumed to be similar to final non-contacts, the latter to final refusers. Although there is evidence that reluctant respondents are not necessarily similar to final nonrespondents, 'difficult' respondents are an interesting group if only because substantial sums of money can be spent in order to reach or convert them. This presentation will discuss different indicators of difficulty and their measurement problems.

Measuring contactability is easier than measuring reluctance although by no means simple. In face-to-face and telephone surveys ease of contact is often measured by the total number of calls made to an address. This is an imperfect measure of response propensity as it may both reflect difficulty of contact and reluctance, requiring further visits for refusal conversion. A better measure is the number of calls to first contact which comprises primary contact attempts only and excludes additional calls to contacted households that should be chalked up to reluctance to cooperate rather than being hard to reach. Neither is the number of calls to first contact a perfect measure in face-to-face surveys, however,

as the timing of calls is not randomly assigned but will be based on local knowledge of the interviewer or on information from previous calls. In addition, persons may be difficult to contact because they are away for a prolonged period (due to a stay abroad or in the hospital) and because they are not at home at the usual times (night nurses, persons who are very active in voluntary work). In the former case a large number of calls within a limited time period will be of no avail.

Studying reluctance is far more difficult than measuring contactability. Reluctance is a multi-faceted phenomenon that can depend on the particular interviewer, the topic of the survey, the perceived response burden or the sponsor, factors that are irrelevant when trying to catch someone at home. Indicators of reluctance or the opposite, willingness to cooperate, sometimes take these substantive differences into account and sometime do not. The simplest indicator of reluctance is whether the respondent did cooperate immediately (including an appointment for an interview) or more visits were required. More complicated indicators measure the number of unsuccessful contacts with the respondent before obtaining cooperation, the type of noncooperation (not able, not willing) and the reason of the initial refusal (no time, wrong subject, surveys not useful, surveys not pleasant). Another distinction is between nonresponse due to situational factors and for more permanent reasons, or between active and passive nonrespondents. Additional indicators of reluctance would be interviewer assessment of future cooperation, and the deployment of additional incentives to convert initial refusers.

Data Quality in the European Social Survey: the Case of Item-nonresponse

Michael Blohm; Centre for Survey Research and Methodology (GESIS-ZUMA), Germany

Achim Koch; Centre for Survey Research and Methodology (GESIS-ZUMA), Germany

The European Social Survey (ESS) is a new multi-country face-to-face survey covering a broad range of social attitudes. Up to now, the survey has been fielded three times (in the years 2002/03, 2004/05, and 2006/07). In each Round, more than 20 European countries participated. The principal aim of the ESS is to measure, monitor and interpret changing public attitudes within Europe. In addition to this substantial aim, the ESS also seeks to advance the methods of cross-national quantitative measurement. The ESS tries to achieve an optimal quality and comparability of its data by specifying targets and survey procedures and by closely monitoring all survey processes. In our presentation we investigate the quality of the collected data. We concentrate on one important aspect of survey data quality, namely item-nonresponse. We analyse whether there are cross-national variations in the level of item-nonresponse rates and try to find out about the reasons for differences by investigating respondent characteristics and the role of the interviewers as determinants of item-nonresponse.

Effects of Survey Sponsorship and Mode of Administration on Respondents' Answers About Their Racial Attitudes

Volker Stocké; University of Mannheim, Germany

This paper analyses whether or not the type of survey sponsor affects respondents' answers about their attitudes toward ethnic minorities, and if so, which processes might be responsible for this effect. This is done with a split ballot experiment where the survey sponsor was either specified to be a university or a commercial marketing research firm. It was assumed that respondents perceive positive racial attitude answers to be more socially desirable when they are interviewed by an academic sponsor and that they adapt their response behavior accordingly. One explanation for such sponsor-induced effects is that they are based on impression-management strategies and are entirely interviewer-mediated in the sense that subjects assume interviewers to share the desirability judgements of the respective sponsor organization. This is tested by randomly assigning subjects to either an interviewer- or a self-administrated mode of data collection, whereby the interviewers' ability to perceive and sanction the responses is systematically varied. Respondents from a random probability sample (N=218) answered the items of the blatant- and subtle-prejudice scales after being randomly assigned to one of the four experimental conditions. As expected, university sponsorship provoked more positive racial attitude answers. However, these

differences were only found for interviewer- but not for self-administered interviews. Starting from an identical average response behaviour under the condition of high response privacy, the university sponsor provoked more positive and the commercial sponsor more negative attitude reports when interviewers recorded the answers. This interaction effect between the type of sponsor and the privacy conditions proved to be statistically significant for the responses on the blatant- but not for the subtle-prejudice scale. In summary, different types of sponsors seem to influence the expectations respondents perceive on the part of the interviewers and thereby subjects' beliefs about which answers are situationally instrumental in order to obtain social approval.

Collecting and Providing Proper Documentation for International Surveys: the Example of the ISSP

Markus Quandt; University of Köln, Germany

The role of organisational aspects in administering surveys 'in the field' is well accepted, since their implications for data quality are rather obvious. Practical organisational tasks like monitoring interviewers, controlling sample selection, or reducing unit non-response are regularly covered in methods textbooks and have been the object of methodological research. Cross-national surveys introduce additional methodological problems, in particular the issue of comparability. However, there is almost no literature on the role of organisational aspects for handling comparability problems on the cross-national level. Systematic research into the question of satisfactory organisational processes is inherently difficult here, because the number of cross-national survey programs is still rather small while the number of relevant factors is large. The presentation will therefore approach that question by discussing the example of specific harmonisation and documentation requirements in the International Social Survey Programme (ISSP). Certain recently adopted procedural changes concern the documentation of socio-demographic variables and illustrate some basic organisational issues very well. The presentation will not recommend an optimal organisational approach to such documentation issues, but it will outline the range of choices to be made.

Data Editing and Quality Control: Value for Money or Art for Art's Sake? the Polish Experience with the ESS

Zbyszek Sawiński; Polish Academy of Sciences, Poland

The presentation addresses a problem of the effectiveness of data editing and quality control procedures applied to survey research results. The presentation starts with the overview of the ESS data processing in Poland to distinguish types of data involved in data editing procedures and to present a sequence in which data are entered and analysed. Special attention is given to three types of questionnaire data: answers to questions asked twice, questions in mutual logical relationships, and data which can be directly compared with external sources. A problem of validity of respondents' answers is discussed to distinguish between data which can be corrected during the editing process and those which should not be corrected. As an example of a practical solution addressing the data quality problem, a transposition procedure is presented, which converts interviewers' reports on visiting respondents (so-called 'Contact Forms') into interviewers' fieldwork histories. The converted data provide an insight into quality of interviewers' work and enable to identify 'strange' interviews which should be verified in the field. At the last step of the presentation, the raw data from questionnaires are compared with the final data file to estimate how strongly the data editing procedures modify original questionnaire recordings. The conclusions are about which methods of data editing are critical for data quality improvement.

Clean Data or Cleaned Data? Data Editing Procedures and Experiences of the ESS Data Archive

Knut Kalgraff Skjåk; Norwegian Social Science Data Services (NSD), Norway

Data editing by the data archives is often considered as the last stage of the survey process before the data are released in workable analysis and documentation tools. The tasks identified in this stage of the survey process are typically to standardise item missing values, to correct typographical errors or out-of-range entries done by punchers or interviewers and to check, and eventually correct, for logical and routing inconsistencies. Furthermore, this final post-data-collection processing often involves imputation and production of different weights and correction coefficients. At last, the data editing comprises of the documentation of data, from assigning names, labels and value labels to the variables, to producing codebooks and comprehensive, computerised meta data documents.

The aim of the archives' editing work is to provide the users with data that are as clean, complete and user-friendly as possible. However, as important this editing process might be, its contribution with respect to reducing survey errors could be limited. Human and technical errors occur in all stages of the survey process, for example programming mistakes, keypunch errors, interviewer errors and misclassifications, and the best way of reducing survey error is to identify and correct as many errors as possible in every single stage of the survey process, in addition to the final processing at the archive.

Comparative surveys like the European Social Survey represent extra challenges in this respect. Time, resources and skills allocated to data control and data editing differ from country to country. Furthermore, data collectors and data producers have different traditions of processing data. Survey organisations typically involved in surveys whose overall task is to produce macroeconomic estimates and time series will have different expectations, procedures and tools than organisations mainly involved in academically driven attitude and value surveys or opinion polls on behalf of the media. In cross-national surveys the data deposited to the archive are therefore not homogenous with respect to control and editing, with implications for true comparative assessments of the survey quality and even directly or indirectly the analysis of the survey results. Ideally, in cross-national surveys the editing activities should therefore be integrated both vertically (survey stages) and horizontally (between countries).

In this presentation the ESS Data Archive's standardised editing procedures will be presented, as well as some experiences from the editing of the two first rounds of the ESS. The possibilities of the integration of data control and editing will be discussed.

A Method for Appending Annual Surveys

Jacopo Pasquini; Health Regional Agency of Tuscany, Italy

Rosa Gini; Health Regional Agency of Tuscany, Italy

Roberto Berni; Health Regional Agency of Tuscany, Italy

Sara Benocci; Health Regional Agency of Tuscany, Italy

Silvia Forni; Health Regional Agency of Tuscany, Italy

Francesca Collini; Health Regional Agency of Tuscany, Italy

Fabio Voller; Health Regional Agency of Tuscany, Italy

Eva Buiatti; Health Regional Agency of Tuscany, Italy

Introduction

The survey "Aspects of daily life" or "Multiscopo", performed since 1993 by ISTAT (Italian Statistical Institute), through direct interview and a self-compiled questionnaire, gathers data on several aspects of ordinary life: job, family, housing, political and social participation, health and lifestyles. Each year the sample includes approximately 24.000 families (54.000 individuals) from 849 different Municipalities of different demographic size. "Multiscopo" has been changing every year adopting new questions on new arguments and adapting itself to international standards. Semantic modifications, concerning questions and proposed answers, have introduced some problems for researchers willing to analyze trends.

Our aim was to define an automatic procedure able to append the annual data while dealing with the changes observed every year in the questions and in the closed answers. The next step was a web based system capable to offer exploration, interrogation and documentation on the generated database.

Methods

Annual codebooks are read by a procedure that traces changes, and generates the code to:

- import the data in the Database Management System (DBMS)
- generate the navigation interface/selection
- generate the necessary documentation of each variable (comments, list of possible answers...)
- manage the connection (importation) of the data in the analysis software

Such procedures demand the participation of the researcher only to estimate the meaning of different questions that can have the same significance in order to establish a trend ("How old are you?" vs "Year of birth"). The system return to the researcher these choices together with the new and the dismissed questions easily allowing any decision on grouping or not the uncertain cases.

Results

The procedures generate two output: the code to append the data and the code to generate the web pages to select and explore the data. Whenever a question (or closed answers) has been modified comparing to the previous year, the system generates two variables: the first one identical to the new question and the second one capable to construct a time trend (in case of closed answer uses the least common multiple of the available answers). The navigation page proposes a table where rows sort all the variables and columns represent the year of each survey. The selection of desired variables is possible by clicking the appropriate checkbox and then a button generates the code to import or connect the data from the DBMS into the desired statistical package. We used Stata and Oracle and obtained more than 1000 variables over 13 years.

Conclusions

All the adopted procedures are independent from the software used and the same system can easily be replicated on other DBMS (MySQL, PostgreSQL). and more codes to import data in other packages (SAS, SPSS, R) can coexist on the same interface. Even if the "Multiscopo" was not born in order to supply time series, and even if this habit arises some methodological questions, this use is very widespread. This method, supplying one instrument for the interrogation, the management of the documentation and the sharing of the data, can also be used for other surveys.

Bazaar Style Data and Metadata in the Age of the Data Cube: a Few Ideas Based on the European Social Survey (ESS)

Hilde Orten; Norwegian Social Science Data Services (NSD), Norway

Over the last decades, the idea of data sharing has been accompanied by a focus on metadata. International standards for describing social science data (eg. the Data Documentation Initiative (DDI)) were developed. The development of software for online data publishing (SDA, NESSTAR etc.) created a new era for the dissemination of integrated data and metadata. Data and metadata were brought closer to each other.

With the latest developments within documentation standards, the distinction between data and metadata erodes. Any piece of information could, at least in principle, be analysable. This logic may be described by the image of a data cube, containing analysable information from a variety of sources. Based on the European Social Survey (ESS), an international project that has put forward an extensive collection of data and metadata, we will present a few ideas about types of analysable information that may be included in the data cube. In a Bazaar style manner, we'll ask our end-users and co-data producers to contribute, by giving their feedback on what they may find useful and what they may provide.

On the Importance of Meta-data and Documentation of Cross-national Surveys

Peter Ph Mohler; Centre for Survey Research and Methodology (GESIS-ZUMA), Germany

Beth Ellen Pennell; University of Michigan, United States

Some years ago, a debate arose in a meeting of the German General Social Survey (ALLBUS) Board about the use of the phrase 'working class' as a descriptive category (see American General Social Survey <http://webapp.icpsr.umich.edu/GSS/> - Variables CLASS and CLASSY). Previous research had shown that marked differences could be obtained in the results of surveys with scales that included or excluded the phrase 'working class.' (Argyle, 1994) The debate became somewhat heated and went back and forth without much progress until Franz-Urban Pappi finally intervened. He pointed out that the 'working-class' category is relevant for those researchers who are interested in the sociological concept of the 'working class'. These researchers do not use such a scale as a standard subjective status self-classification, but as a specific tool to investigate respondents' self-classification as 'belonging to the working class as a societal concept'. The debate ceased immediately and there was consensus that this difference should have been documented in the papers presented to the board.

This debate gets to the heart of the purpose of this paper: the need of relevant, comprehensive and easily retrievable documentation of the entire survey production process from questionnaire to data file. While this is a complex endeavour for single-country studies such as the ALLBUS, the challenges increase exponential when one adds 2, 3, 4... 20... 30 and more countries.

The paper will shortly discuss the interplay between numerical data, meta-data and para-data. It will concentrate on the issue of achieving concepts of relevant, cascaded and modular meta-data documentation and show how necessary comprehensive tools are for cross-national surveys, which allow data producers to provide good meta-data for multiple users such as themselves, data archives and analysts.

6.3 The Use of Paradata for Evaluating and Improving Survey Quality

Session coordinated by:

Annette Jäckle; University of Essex, United Kingdom

Peter Lynn; University of Essex, United Kingdom

Frauke Kreuter; University of Maryland, United States

Presentations

1. *Cooperation in Centralised CATI Panels - a Contact Based Multilevel Analysis to Examine Interviewer, Respondent, and Contact Effects* – O. Lipps (Switzerland)
2. *Explaining Differences in Contact and Co-operation Rates Across Countries* – A. Blom (Germany), A. Jäckle (United Kingdom), P. Lynn (United Kingdom)
3. *An Application of Sequence Analysis to Contact Data* – F. Kreuter (United States), U. Kohler (Germany)
4. *Result of a Nonresponse Follow Up Study in a Multi-mode National Survey* – R. Schnell (Germany)
5. *Using Paradata to Improve Nonresponse Adjustment* – C. Casas-Cordero (United States), R. Groves (United States), F. Kreuter (United States), M. Lemay (United States), K. Olson (United States), A. Peytchev (United States), T. Ragunathan (United States), J. Wagner (United States), T. Yan (United States)
6. *A Dynamic Approach to Survey Participation* – K. Olson (United States)
7. *Active Management – Survey Monitoring* – F. Laflamme (Canada), M. Maydan (Canada)
8. *Automated Interviewer Call Records: a Tool for Improved Survey Management in a Large Survey Organisation.* – G. Nicolaas (United Kingdom)
9. *Building Dynamic Survey Cost Models Using Survey Paradata* – R. Groves (United States), E. Peytcheva (United States)

Presentations in the session “The Use of Paradata for Evaluating and Improving Survey Quality”

Cooperation in Centralised CATI Panels - a Contact Based Multilevel Analysis to Examine Interviewer, Respondent, and Contact Effects

Oliver Lipps; Swiss Household Panel, Switzerland

Using call records from two Swiss Household Panel surveys, we analyse contact performance on final cooperation. We distinguish between contacts on household reference persons, having to complete the household grid questionnaire, and “other” eligible household members, having to complete their individual questionnaires.

As to the first contact, the interviewer specific variance amounts to 5% for the reference person and 2% for the other persons. In higher order contacts the interviewer effects are much smaller. Socio-demography and attitudes are more important role for the reference persons. The contact number has negative effects on contact performance for the reference person, positive for other persons. These findings call for a quick realisation of the household grid questionnaire, possibly by a reassignment of the best interviewers to conducting the first contact. Interviewer learning effects appear to be relevant only in a negative sense, i.e. while having experienced refusals seems to have negative effects on the next contact outcome, having experienced completed interviews does not seem to have any effects. Using the same interviewer does not have positive effects.

Explaining Differences in Contact and Co-operation Rates Across Countries

Annelies Blom; Centre for Survey Research and Methodology (GESIS-ZUMA), Germany

Annette Jäckle; University of Essex, United Kingdom

Peter Lynn; University of Essex, United Kingdom

Inferences about differences between subgroups rely on the assumption that survey errors are comparable across groups. If this assumption fails and sampling or non-sampling errors (of coverage, non-response or measurement) differ, then any differences detected between groups may merely be artefacts of the data. This problem arises for any group comparison, but is accentuated for cross-national comparisons, where the scope for differential errors between groups is multiplied. Cross-national studies vary in their efforts to ensure data comparability, with the ‘strictest’ studies enforcing standardised data collection procedures across countries. The same procedures need, however, not necessarily lead to the same outcomes. We use call record data from the first wave of the European Social Survey to examine the determinants of contact and co-operation across countries.

Three factors may explain differences in outcomes across countries: 1) differences in population characteristics that are predictive of contact or co-operation, 2) differences in fieldwork procedures that influence the outcome, 3) differences in at home patterns and response behaviours of sample persons (model coefficients) or differences in unobserved factors affecting the survey taking climate (model residuals). We disentangle how much of the differences in outcomes are due to differences in each of these factors.

The findings inform the design of optimal fieldwork procedures, by shedding light on the extent to which the actual procedures were comparable across countries and the extent to which idealised comparable procedures would yield equivalent outcomes. If outcomes are explained by differences in manipulable fieldwork characteristics, then differential errors related to non-contact and non-response can be reduced by standardising procedures across countries. If non-contact and non-response are explained by interviewer or sample person behaviours, then standardising fieldwork procedures cannot ensure comparable outcomes. The findings also inform strategies for adjusting for unit non-response in cross-national studies, by examining whether biases associated with non-contact and non-response differ across countries. If outcomes are explained by observable population characteristics, then the

same adjustment models can be used cross-nationally. If outcomes are explained by differences in behaviours of sample persons and interviewers, then adjustment models need to be country specific.

An Application of Sequence Analysis to Contact Data

Frauke Kreuter; University of Maryland, United States

Ulrich Kohler; Wissenschaftszentrum Berlin für Sozialforschung (WZB), Germany

More and more survey organizations collect contact history data during the data collection process. These data usually include day and time of contact, mode in which contact was attempted, and outcome of the contact. Sometimes interviewer observations are made during the contact attempt and reasons for refusal are noted. Contact and other process data can guide mid-survey decisions that affect nonresponse, measurement, and sampling variance properties of the resulting statistics such as with responsive survey designs. In addition to the possibility for on-going design decisions, contact data might be useful in creating nonresponse adjustment weights. For example, conditional on the use of a balanced call schedule, the information available from contact attempts can be used to create estimates for “time spent at home”. “Time spent at home” is a predictor of the probability of making a successful final contact given a fixed field period. “Time spent at home” can likewise be highly correlated with substantive outcome variables such as intra-household provision for after-school childcare, hours of TV viewing, or employment status. A nonresponse weighting procedure that takes the number of contact attempts into account can therefore reduce bias for particular outcome variables. The mere count of contact attempts might however not reveal the whole story of a contact, and other summary measures of the contact history might be more informative for respondent behaviour.

This talk will evaluate the extent to which sequence analysis can effectively describe and analyze contact data by taking the sequence of events into account. Sequence analysis is most often used in genetic applications where entities (DNA) are built by a limited number of elements (bases) that are ordered in a specific way. Other examples for sequences are songs that are built by tones that appear in a specific order, or careers of employers that are built by specific job-positions and ordered along time. For contact protocol data, we will demonstrate tools for describing and visualizing sequences as well as examples on how to perform optimal matching to form clusters of similar sequences. Our main interest is on describing and grouping respondent behaviour. However, certain patterns may suggest changes in specific interviewer behavior or organizational practices. The usability of sequence analysis summary statistics for nonresponse adjustments will be discussed.

Result of a Nonresponse Follow Up Study in a Multi-mode National Survey

Rainer Schnell; University of Konstanz, Germany

In 2000, we conducted 4 national fear of crime survey using 160 PSUs with an interpenetrated design in 3 response modes (F2F, CATI, Mail). For each of the 4 surveys, a thorough nonresponse follow-up by CATI has been done. Furthermore, for each selected respondent a huge number of variables concerning housing, fear-of-crime-related variables like protective measures and the social environment has been recorded. By this design we can compare the effect of different nonresponse correction procedures on the estimates with the results of the nonresponse-follow-up. On descriptive statistics and multivariate models, we found strong similarities between the MI and the follow-up-estimates by comparing naive estimates, various weighted estimates, estimates after multiple imputation with the follow-up-estimates. This may suggest, that if the assumptions of MAR hold, carefully selected and collected additional data applied in a MI could yield similar estimates to a nonresponse follow-up.

Using Paradata to Improve Nonresponse Adjustment

Carolina Casas-Cordero; University of Michigan, United States

Robert M. Groves; University of Michigan, United States

Frauke Kreuter; University of Maryland, United States

Michael Lemay; University of Michigan, United States

Kristen Olson; University of Michigan, United States

Andy Peytchev; RTI International, United States

Trivellore Ragunathan; University of Michigan, United States

James Wagner; University of Michigan, United States

Ting Yan; University of Michigan, United States

Post-survey weighting adjustments are a widely used method for addressing survey nonresponse. To be effective, variables used for nonresponse adjustment need to be highly correlated with both response propensity and with the survey variable of interest. Propensity models and weighting class adjustments that are commonly used to create nonresponse adjustment weights employ easily available variables such as gender, sex, and race. Unfortunately, these variables are not always closely related to response probability, and may be only loosely correlated with the outcome variables of interest. However, interviewers can make observations about the sample member, the housing unit, and the doorstep interaction that could be related not just to the propensity to respond, but also to statistics of interest. In addition call record data usually collected for survey management purposes may plausibly be related to participation propensities and outcome variables. For example the information available from contact attempts can be used to create estimates for "time spent at home". "Time spent at home" is a predictor of the probability of making a successful final contact given a fixed field period. "Time spent at home" can likewise be highly correlated with the substantive outcome variable such as intra-household provision for after-school care for children.

This paper will show several examples for the use of paradata in nonresponse adjustments for a set of surveys that differ in topic, location, and survey organization conducting it. Those surveys are the National Survey of Family Health (NSFG), the National Election Study (NES), the Consumer Expenditure Survey (CES), the Los Angeles Family and Neighborhood Survey (L.A.FANS), the University of Michigan Transportation Survey (UMTRI) as well as the European Social Survey (ESS).

A Dynamic Approach to Survey Participation

Kristen Olson; University of Michigan, United States

Traditional views of survey participation view a person's likelihood of participating in a survey as either static or stochastic. Both views assert that a sample unit has one response propensity, given an evolved survey protocol. In a field survey, sample units vary in their receipt of individual recruitment protocol components and in the timing of these components. These components are specifically applied to alter response propensities. This paper argues that sample units have more than one response propensity, changing with the introduction of new protocol components. We call this a dynamic view of response propensity. Dynamic response propensities are estimated using two surveys with records for respondents and nonrespondents. The records are used to illustrate how a dynamic response propensity approach changes the understanding of nonresponse bias and of recruitment procedures.

Active Management – Survey Monitoring

François Laflamme; Statistics Canada, Canada

Mike Maydan; Statistics Canada, Canada

With the increasing complexity of survey design, survey managers need better tools to assess the data collection survey process. The development of data collection computer-assisted methods has opened the door to a wide scope of process data ("paradata") that can be used to evaluate the data collection survey process status in a timely manner.

Active Management at Statistics Canada is a set of plans and tools developed to manage survey data collection while it is in progress. This includes monitoring progress (e.g. response rates, survey progress, costs, data quality, etc...), timely analysis of these indicators, identifying problems, implementing and communicating corrective actions, and evaluating success. The two main objectives of Active Management are to identify problems as early as possible, and correct them before collection has finished and to use collection resources effectively, in a way that optimizes data quality which refers to responsive design strategy.

The paper begins with an introduction to the key components of the Active Management including the description of the standard and customized reports. The second part presents a series of practical example of customized reports currently used to monitor issues and data quality indicators as well as the current status of the corrective and communication action plan in Statistics Canada.

Automated Interviewer Call Records: a Tool for Improved Survey Management in a Large Survey Organisation.

Gerry Nicolaas; National Centre for Social Research - NatCen, United Kingdom

Over the years, data collection operations at the National Centre for Social Research (NatCen) have become larger and more complex. In order to manage the data collection process more effectively, NatCen has developed a custom-built automated system for collecting interviewer call records. This system was successfully implemented in January 2006 for all of its CAPI surveys.

The system records information such as postcode, date and time of each call, outcome of each call, interviewer number, and project number. The call data can be linked to other data sources, including area data (e.g. Census), information about the interviewer (e.g. tenure, grade), interviewer observations and questionnaire data. The strength of these data is that they cover all CAPI surveys in the field over any period of time. Working on more than one survey is common within large organisations such as NatCen, and is therefore important in terms of planning and monitoring interviewer behaviour.

The focus of this presentation will be on the use of these records to improve survey management. We will consider how these data can be used to plan fieldwork effectively, to monitor and supervise interviewers in the field, and to inform decisions about cost and quality trade-offs either at the design stage of a survey or during data collection (i.e. responsive survey design). We will also address some of the challenges involved in using these data for survey management.

Building Dynamic Survey Cost Models Using Survey Paradata

Robert M. Groves; University of Michigan, United States

Emilia Peytcheva; University of Michigan, United States

Many household surveys have experienced large increases in costs over the past few years, as repeated callbacks are required to obtain high response rates. Even survey methods that do not use probability sampling methods appear to experience higher costs, as substitution of cases must necessarily follow failed attempts to gain cooperation. By and large, these cost increases have not arisen from design changes that have obvious cost-inflating effects. Instead, changes in the public's at-home patterns, door- and telephone-answering behavior, and tendencies toward reluctance to participate combine to increase the total interviewer effort required to obtain an interview.

These last few years have probably produced increased mis-estimation of survey budgets because of the dynamic nature of the behavioral changes above. When initial survey budgets are ill-conceived, either reduced sample size, lower response rates, or cost overruns occur.

The paper proposes a dynamic production model that is based on an amount of effort applied to a set of active cases, with a productivity adjustment that is a function of the relative difficulty of the remaining set of cases. The empirical specifications of these models attempt to reflect the fact that one hour of interviewer effort has differential impact on the production of interview data records as a function of the lifecycle of a sample administration.

Data to test alternative specification of the production models come from a continuous interviewing survey conducted at the Michigan Survey Research Center, using an area probability sample with face-to-face interviewing. The paradata used to estimate the model arise from interviewers uploading paradata describing their work day to a central server once every 24 hours.

The paper concludes with summaries of how practical survey work can be informed by such cost models, both reduce the threat of cost overruns and to balance costs and errors of survey estimates.

6.4 Using Contact Forms for Data Quality Assessment of Cross-national Surveys

Session coordinated by:

Jaak Billiet; University of Leuven, Belgium

Ineke Stoop; Social and Cultural Planning office (SCP), Netherlands

Presentations

1. *Studying Survey Co-operation in Perspective of Timing During the Contact Procedure: Analysis of Contact Forms Among Selective Countries on European Social Survey Round 2* – H. Matsuo (Belgium), G. Loosveldt (Belgium), J. Billiet (Belgium)
2. *Satisficing for Reluctant Respondents in a Cross-national Context* – O. Kaminska (United States), J. Billiet (Belgium)
3. *Cross-national Differences in Response Rates in the 2nd Round of the European Social Survey: an Assessment and Attempt to Explanation* – L. Vandecasteele (Belgium), J. Billiet (Belgium)

Presentations in the session "Using Contact Forms for Data Quality Assessment of Cross-national Surveys"

Studying Survey Co-operation in Perspective of Timing During the Contact Procedure: Analysis of Contact Forms Among Selective Countries on European Social Survey Round 2

Hideko Matsuo; University of Leuven, Belgium

Geert Loosveldt; University of Leuven, Belgium

Jaak Billiet; University of Leuven, Belgium

This paper consists of 2 different parts. In the first part, it studies survey cooperation focusing on the timing of different aspects of contact procedure (occurrence and timing of event; duration of events) by making use of standardized contact forms and applying life table method. We make use of all information pertaining to the occurrence and timing of each visits and its outcome that is recorded up until the 10th visit. On the basis of this, first contact outcome and final contact outcome (occurrence, timing and outcome) is obtained to study trends on survey cooperation to map the cumulative distribution of total contact procedure and lengths (mean duration) in terms of number of days per country. Trends of cumulative distribution are also presented in relation to socio-demographic variables (age, gender and region if available). On the basis of the following indicators at the country level:

1. total response rate;
2. average number of contact attempts and effective contacts (face to face contacts with the respondent); and
3. existence of large-scale refusal conversion activities, the aforementioned analysis is proposed for 4 selective countries (ex. Belgium, Netherlands, Greece, Germany or Switzerland) of Round 2.

In the second part, in-depth analysis is conducted on the event of survey cooperation by including number of covariates, for instance, different types of contact variables (contact attempts and successive effective contacts) and fieldwork variables (load of interviewer and change of interviewer or not) by controlling number of socio-demographic variables (age, gender and region). From the analysis above, 2 countries are considered for this analysis. Cox regression model is applied, to study the effect of these numbers of independent variables on the timing to event (dependent variable). Three different sub-groups are studied within the same context of event of survey cooperation:

1. respondent vs non-response group;
2. converted respondent vs. co-operative respondent;
3. converted respondent vs remaining refusers.

Based on our analysis, our paper tries to propose recommendations with regard to the fieldwork guidelines and also discusses the indication of our results to non-response biases.

Satisficing for Reluctant Respondents in a Cross-national Context

Olena Kaminska; University of Nebraska, Lincoln, United States

Jaak Billiet; University of Leuven, Belgium

Reluctant respondents are likely to have lower motivation to answer survey questions. Lack of motivation for converted reluctant respondents can lead to using heuristic shortcuts when processing the answer as opposed to thinking through the answer carefully. This can result in lower answer quality for reluctant respondents in comparison to cooperative respondents. A potential risk in cross national surveys occurs when different nationalities have different proportion of cooperative respondents or when efforts for

converting reluctant respondents differ across nations either because of differences in culture of survey organizations or because of budget differences. As a result different answer quality can be observed across countries.

The purpose of our study is to investigate whether reluctant respondents are more likely than cooperative respondents to provide satisficing responses and whether this pattern differs across nations. The study is based on the data from European Social Survey (ESS), round 1, for Netherlands and Germany, the two countries that have more than 450 converted refusals. Three categories of reluctance to respond are created using information from contact forms: cooperative respondents, soft and hard refusals. Satisficing is assessed using two types of measures. First measure is based on interviewer evaluation of respondent's effort in answering the questions. Alternatively, following Krosnick (1991), four measures of satisficing are created: don't know responses, straight responses, acquiescence (agreeing) and easy responses (including extreme and middle responses).

Our results provide some evidence that satisficing increases with reluctance to respond. There is a clear difference in the pattern between the two countries: according to interviewer evaluation, reluctant respondents in Germany are less likely than reluctant respondents in Netherlands to provide satisficing answers. Similar pattern is observed for alternative measures of satisficing: controlling for other potential confounding factors, such as age, education and gender, we find that while reluctant respondents in Netherlands are more likely to provide straight and easy responses, this is not true for reluctant respondents in Germany. The results and their potential impact for cross-national surveys are discussed.

Keywords: reluctant respondent, soft and hard refusal, satisficing, straight responses, acquiescence, extreme responses, don't know responses, data quality, cross-national survey.

Cross-national Differences in Response Rates in the 2nd Round of the European Social Survey: an Assessment and Attempt to Explanation

Leen Vandecasteele; University of Leuven, Belgium

Jaak Billiet; University of Leuven, Belgium

Despite the intended harmonisation in survey design and organisation, the European Social Survey (ESS) shows substantial country differences in response, refusal and non-contact rates. Yet, also considerable country differences in survey procedures prevail (See: Billiet and Phillippens, 2004). In the first part of this paper, an explanation for differing response rates will be sought in country differences related to contact mode, timing and number of contacts. In the second part of the paper, a multilevel logistic regression model is presented in which the effect of survey procedure characteristics is investigated next to respondent-specific variables. The latter model gives an indication of the relative importance of several survey procedures on the response pattern, as well as the bias created by non-response in certain population groups. But definitely as interesting is a discussion of remaining response differences between the countries, once person-specific characteristics and survey procedures have been controlled for. These will be clarified on the basis of indicators of the survey climate in the country.

6.5 Mobile Phone Usage in Survey Research: Implications for Data Quality

Session coordinated by:

Mario Callegaro; University of Nebraska, Lincoln, United States

Marek Fuchs; University of Kassel, Germany

Presentations

1. *Profiling Different Types of Telephone Households: Who Are Researchers Potentially Missing?* – M. Galin (United States), V. Bosch (Germany), J. Sauer (Germany)
2. *Mobile Phone Surveys in Germany* – S. Häder (Germany), S. Gabler (Germany), G. Schneiderat (Germany)
3. *Feasibility of Conducting Cell Phone Samples* – J. Dayton (United States), R. ZuWallack (United States)
4. *Ring Ring: the Possibility to Complement a Swedish Fixed Line RDD Sample with Mobile Phone-only Users* – U. Isander (Sweden), A. Jaktlund (Sweden)
5. *Combining Cell Phone and Landline Samples: Dual Frame Approaches* – R. ZuWallack (United States)
6. *Pilot Development of a Smartphone-enabled Full-probability Panel* – C. Hill (United States), P. Biemer (United States), J. Eyerman (United States), A. Peytchev (United States), D. Roe (United States)

Presentations in the session "Mobile Phone Usage in Survey Research: Implications for Data Quality"

Profiling Different Types of Telephone Households: Who Are Researchers Potentially Missing?

Mickey Galin; Mediamark Research, United States

Volker Bosch; GfK Aktiengesellschaft, Germany

Juliane Sauer; GfK Aktiengesellschaft, Germany

The spread of cellular phones in the United States has realized persistent growth in the last several years. As of 2006, Mediamark Research found that 83.46% of American households had at least one cell phone and 11.66% of American households had no landline but at least one cell phone (MRI Fall 2006 Study). Several government studies support these data (the Census Bureau's Current Population Survey and the National Health Interview Survey).

This proliferation of cell phones and particularly the growth in cell phone only households have implications for telephone based survey research. An article in the spring 2007 issue of Public Opinion Quarterly even explores issues of contacting and surveying cell phone only populations versus cell phone and landline households in the United States. A prominent, up-coming research conference will even devote a number of sessions to this topic. At the same time, cellular telephone proliferation and its implication for survey research is not exclusive to the United States.

This paper seeks to present unique data collected by Mediamark Research for its National Survey of the Consumer Study to profile various telephone populations (landline only, both landline and cell phone, cell phone only). Through demographic, media consumption and product consumption data, these different populations will be discussed and compared. We will address the question: by missing certain populations in a given study, does one introduce systematic bias to that study?

We will also introduce a more international perspective by including some similar information from several countries in Europe (i.e. the United Kingdom, Slovakia and Germany).

Mobile Phone Surveys in Germany

Sabine Häder; Centre for Survey Research and Methodology (GESIS-ZUMA), Germany

Siegfried Gabler; Centre for Survey Research and Methodology (GESIS-ZUMA), Germany

Götz Schneiderat; University of Dresden, Germany

The number of households solely equipped with fixed phone lines has decreased in Germany since 2003. Simultaneously the number of people using mobile phones exclusively has increased; especially among young males living in single households. Latest estimations indicate that 7- 8 percent of households have mobile telephones only. Experiences from other countries are similar and, in addition, show that the speed at which fixed phone lines are replaced with mobile phones will even increase in the future. This change in telephone coverage cannot be ignored in survey practice, since current frames for the selection of households include fixed line phones only.

To shed light on this issue the German Research Foundation is funding a joint research project of the TU Dresden and the Centre for Survey Research and Methodology, Mannheim. The project aims to develop possibilities for combining fixed line and mobile phone surveys.

The main issues are: - developing a sampling frame for the selection of mobile phone numbers

- conducting a parallel survey using a sample of fixed line phones (n=1500), as well as one of mobile phones (n=1500)
- developing a weighting model for combining the two samples.

In the paper we will explain the outline of the project and present the results of two pretests.

Feasibility of Conducting Cell Phone Samples

James J. Dayton; ORC Macro, United States

Randal S. ZuWallack; ORC Macro, United States

In response to growing concerns about the inadequate frame coverage of traditional random digit dialing (RDD) samples, which exclude wireless telephones, Macro International conducted over 1900 interviews in six partnering states: Connecticut, Florida, Massachusetts, Montana, New Jersey, and Texas. The interviewing spanned October 2006 through February 2007. This cell phone pilot study was initiated to advance our understanding about cell phone sample surveys, the operational and financial feasibility of conducting cell phone samples to supplement traditional RDD samples, and the consequences of excluding the cell phone only population from survey estimates.

The general methodology for conducting the interviews was similar across the states. The one variation was that Massachusetts offered incentives to respondents upon completing the survey to help offset any consumer costs for inbound calling. There were some variations in the survey content. A general health survey was administered in each state with questioning on health topics such as general health and exercise, health access, diabetes, asthma, tobacco use, and alcohol consumption. We also ask questions about demographics and cell phone usage. Florida and New Jersey included questioning on smoking cessation, while Connecticut, Massachusetts, Montana, and Texas included questioning on women's health, HIV/AIDS, seatbelt use and drinking and driving.

Without list-assisted luxuries for sampling cell phones and imposed legal restrictions on the use of predictive dialers to call cell phones, we experimented with other methods to build in sampling efficiencies. In each state, we experimented with a two stage cluster sample inspired by Mitofsky-Waksberg methods, where we use working cell phone numbers to identify clusters where we draw second stage units.

This presentation discusses our general experiences with conducting cell phone surveys in these six states. We discuss usage statistics, evaluate the current feasibility of supplementing traditional RDD surveys with cell phone interviewing, and discuss the results of the Mitofsky-Waksberg experiment.

Ring Ring: the Possibility to Complement a Swedish Fixed Line RDD Sample with Mobile Phone-only Users

Ulf Isander; Sifo Research International, Sweden

Asa Jaktlund; Sifo Research International, Sweden

SIFO Research International's weekly Omnibus Survey employs a sample of 1,000,000 telephone numbers that are drawn from the national telephone directory of private fixed line subscribers. The Omnibus is used for conducting pre-election political polls among various other consumer surveys. In Sweden, the use of traditional fixed line telephony has been, and still is very widespread. 89% of households have a fixed line subscription. (National Post and Telecom Agency, -07). Due to this fact, the Omnibus method of sampling has worked well so far. However, there is increasing worry that the omission of mobile only subscribers could cause bias to the results although this part of the population is still small in relation to other European countries (Kuusela et al, -07).

Several other studies in the U.S and Europe have discussed issues concerning the impact of the growing number of mobile phone only users (Keeter, -06, Callegaro & Poggio, -04, Ve-hovar et al., -04, Kuusela et al, -07). However, no such study has yet to our knowledge been conducted in Sweden. Following this, in order to try to obtain information on the Swedish situation we attempted to sample mobile-phone-only users and compare their answers to those from the standard Omnibus sample. During February 2007 we conducted telephone interviews with questions on socio-demographic variables, values and political preferences. The same questions were also included in the Omnibus during the same time period for comparisons. We used a sample of 1067 individuals without registered fixed telephone numbers. 243 successful interviews were conducted. The paper will discuss issues concerning the possibility of obtaining quality samples and the possibility to compare estimates.

Combining Cell Phone and Landline Samples: Dual Frame Approaches

Randal S. ZuWallack; ORC Macro, United States

As recently as 2003, the percentage of adults living in households without landline telephone service in the United States was estimated by the National Health Interview Survey to be 4.4 percent, with 1.6 percent having no telephone at all and 2.8 percent having only a cell phone (Blumberg and Luke, 2007). Three years later, according to the same report, the percentage of adults in households with no telephones at all has held steady at 1.8 percent, yet the percentage of adults who live in wireless only households has climbed to 9.6 percent. As we witness the technological and cultural shift in telecommunications, the telephone research community is actively pursuing new ideas and methods to keep pace.

As more and more consumers adopt the wireless lifestyle, traditional random digit dialing (RDD) samples will be scrutinized and challenged as not being representative of the target audience. A potential solution to the frame undercoverage is to supplement traditional land-line telephone surveys with surveys conducted with respondents selected via a cell phone sample. There are barriers to overcome—inbound charges and safety concerns—but advantages to harness. As a personal device, cell phones provide a direct communication link to respondents, particularly young adults, at virtually any time of day. Our research suggests that most respondents have their phones on all or most of the day and a high percentage are between the ages of 18 and 24, a group whose share has declined in contemporary telephone samples.

Supplementing a landline sample with a cell phone sample improves the population coverage, but since the frames overlap in their population coverage, proper selection probabilities are crucial to developing an appropriate weighting scheme. We develop a framework for establishing the overlap in frame coverage for combining the cell phone samples with traditional BRFSS samples. We present two methods for weighting and use a simulation to evaluate the methods under various non-response conditions. Finally, coupling results from interviews conducted via cell phone from October 2006 through February 2007 with traditional landline surveys, we examine the application of these weighting methods.

Pilot Development of a Smartphone-enabled Full-probability Panel

Craig A. Hill; RTI International, United States

Paul Biemer; RTI International, United States

Joseph Eyerman; RTI International, United States

Andy Peytchev; RTI International, United States

David Roe; RTI International, United States

This paper will describe a project we are currently undertaking to design, implement, and standardize methodologies for the use of smartphone technology to collect real-time self-reported data from a panel of study participants. The paper covers three ongoing phases of the study: development, recruiting and data collection, and analysis/findings.

Phase I. Development

- Develop a screening and recruiting tool based on a smartphone platform. This tool will be used by the field staff to locate, contact, and recruit sample members into the panel study.
- Develop a real-time survey tool based on the smartphone platform. This tool will be used to notify study participants when it is time to complete a survey, and to collect and transmit self-reported information from the study participants.
- Develop a module for training the field staff on recruiting procedures, software guidelines, and hardware care and use. Develop a second module for training study participants on the data collection procedures, software characteristics, and hardware care and use.

- Develop a case management system for tracking the disposition of each study participants, the status of each administered survey, and the metadata from the interview process. Meta-data can include the time required to complete a survey, the responsiveness of each participant to survey requests, and any systematic problems with the process, software, or hardware.
- Develop panel maintenance methods tailored to this population and this technology in order to retain as many of the recruited participants as possible throughout the study period.

Phase II. Recruiting and data collection

- Screen a sample of adults from the local area and recruit them to be participants in the panel. We expect to screen about 150 households in order to build a panel of 100 participants.
- Implement training modules, panel maintenance methods, and ongoing surveys. We expect to administer 6 monthly surveys of about 20 questions each, beginning in April.

Phase II. Data Analysis/Findings

- Conduct near real-time analysis of the data as they are collected in each of the monthly surveys. These data will be analyzed and used for policy analysis and other research projects.
- Analyze the metadata throughout the study to diagnose and resolve any procedural, software, or hardware problems. Conduct a more detailed analysis of the metadata to identify common problems and relate them to participant characteristics.
- Develop protocols to guide the use of smartphone technologies in future studies.
- Assess the potential of this technology for simple and complex data collection.

6.6 Mixing Modes of Data Collection

Session coordinated by:

Caroline Roberts; City University, United Kingdom

Presentations

1. *Mixing Modes of Data Collection in Surveys* – C. Roberts (United Kingdom)
2. *Evaluating the Impact of Data Collection Mode Upon Response to Subjective Surveys: Main Results from the Mode Arts Systematic Review* – M. Robling (United Kingdom), K. Hood (United Kingdom), D. Ingledew (United Kingdom), I. Russell (United Kingdom), G. Greene (United Kingdom), A. Sayers (United Kingdom), R. Ivans (United Kingdom), L. Sander (United Kingdom), J. Williams (United Kingdom), C. Shaw (United Kingdom)
3. *Mixed Mode Data Collection in Europe* – G. Eva (United Kingdom), S. Widdop (United Kingdom)
4. *Causes of Mode Effects: Separating out Interviewer and Stimulus Effects in Comparisons of Face-to-face and Telephone Surveys* – C. Roberts (United Kingdom), A. Jäckle (United Kingdom), P. Lynn (United Kingdom)
5. *Survey Errors in Mixed-mode Pilots That Incorporate Face-to-face, Telephone, Paper and Web Interviewing* – B. Janssen (Netherlands), B. Schouten (Netherlands)
6. *Measurement Equivalence of Paper-and-pencil and Online Organizational Surveys: a Large Scale Examination in 16 Countries* – A. De Beuckelaer (Belgium), F. Lievens (Belgium)

Presentations in the session "Mixing Modes of Data Collection"

Mixing Modes of Data Collection in Surveys

Caroline Roberts; City University, United Kingdom

In this paper I provide an introduction to the session by presenting an overview of the field of research into mixed mode data collection in surveys. I begin with a discussion of the factors influencing mode choices in surveys and the advantages and disadvantages of different modes with respect to a range of criteria, including their impact on data quality (so-called 'mode effects'). Increasingly, survey designers are exploiting the potential offered by using a combination of data collection modes, either to offset the weaknesses of a particular mode with the strengths of another or, for example, to try to reduce the overall costs of fieldwork. The paper describes the major challenges involved in mixing modes, focusing in particular on the problem of mode effects on measurement error – or how people respond to survey questions. Modes of data collection appear to affect respondents' answers by influencing either a) the amount of effort needed to answer the question; or b) the respondents' willingness to answer questions honestly, leading to a range of response errors including satisficing effects and social desirability bias. Mode effects have implications for the comparability of data collected in different modes; understanding their causes is, therefore, an important step in developing ways of reducing the negative impact of mixing modes on data quality. This review of the literature was conducted in the context of a wider programme of research into the feasibility of mixing modes on the European Social Survey. The paper introduces this programme of research, describing some of particular challenges involved in mixing modes in comparative surveys and in considering a switch in data collection strategy in the context of a time series.

Evaluating the Impact of Data Collection Mode Upon Response to Subjective Surveys: Main Results from the Mode Arts Systematic Review

Mike Robling; Cardiff University, United Kingdom

Kerenza Hood; Cardiff University, United Kingdom

David Ingledew; University of Bangor, United Kingdom

Ian Russell; University of Bangor, United Kingdom

Giles Greene; Cardiff University, United Kingdom

Adrian Sayers; Cardiff University, United Kingdom

Rhys Ivans; Cardiff University, United Kingdom

Lesley Sander; Cardiff University, United Kingdom

John Williams; University of Wales, Swansea, United Kingdom

Chris Shaw; University of Glamorgan, United Kingdom

Background: Technological innovation in telephony and computing has resulted in an increased variety and complexity of options for data collection mode in the health sciences and (for example) the application of unimodal and multi-modal approaches. Whilst pragmatically sensible, such approaches need to be informed by both theoretical models of survey response and, if necessary, statistical consideration of adjustment for biases introduced by multi-mode usage. The MODE ARTS study aims to identify generalisable features affecting responses to different modes of data collection relevant to health research from a systematic review of the literature.

Methods: 19 databases and selected journals were searched covering the fields of health, social sciences, evidence based medicine (e.g. DARE), economics and sports (up to the end of year 2004). The search strategy implemented a matrix defined by types (groups) of data collection mode (e.g. technology assisted, paper-and-pencil administration). Studies were included in the review if they featured comparison of two or modes of data collection for a quantitatively measured subjective outcome.

Article retrieval was not restricted by language or country of origin. A three stage review process was employed with firstly abstract-only review, then initial screen of full paper and finally full data extraction.

Results: 63,305 citations were initially identified, 39,253 of which were unique. Following all review stages, 404 studies remained in the review (covering publications from 1957 to 2004), the majority being North American in origin (61.7%) and in English (76%). Most studies originated from academic (39%) or health (46%) institutions. The proportion of within- (48%) and between-group (52%) studies was similar. The paper will present the main results of the review which is based on three key summary statistics for each comparison between two modes – absolute difference between standardised means, ratio of standardised variances and effect sizes. Multi-level analysis will adjust for clustering of comparisons within studies. Sensitivity analyses will explore the impact of weighting by study quality scores.

Discussion: This systematic review has drawn evidence from a wide range of professional disciplines to identify the impact of data collection mode upon response to subjective survey questions. The review and its analysis has been informed by a cognitive approach to how modes may influence response, most notably that represented by Roger Tourangeau. How the results fit with existing evidence about mode effects will be discussed in the presentation.

Key words: modes of data collection, systematic review, cognitive approaches to survey methodology

Mixed Mode Data Collection in Europe

Gillian Eva; City University, United Kingdom

Sally Widdop; City University, United Kingdom

Cross-national surveys face a range of obstacles to achieving reliable, comparable data and generally try to resolve this by making survey practices (sampling, question-wording, and so on) as similar as possible. For this reason, the European Social Survey (ESS) currently allows only face-to-face data collection, generally considered to be the mode that results in the highest quality data, especially for a long, large-scale, population survey such as the ESS. Yet now, partly for the same reasons that mixed mode data collection is currently becoming increasingly widespread in single-nation studies (e.g. due to declining response rates and increasing fieldwork costs), the possibility of introducing mixed modes on the ESS is being considered.

However, there is currently a lack of empirical data supporting the assumption that mixed mode data collection, or simply a move away from face-to-face data collection, is desired, or indeed possible, on a survey like the ESS. For this reason, a ‘mapping exercise’ is being carried out at City University in order to get a comprehensive picture of current survey practice across Europe. Specifically, this research aims to establish the capacity and demand for mixing modes of data collection in ESS participating countries. The project consists of desk research and a consultation exercise with survey practitioners across Europe, from which we will present some preliminary findings.

This paper begins by outlining the background to the research project and the methods used so far. It then describes some examples of the current use of mixed modes across Europe, looking specifically at what this tells us about the possibility of mixing modes in the ESS. We then discuss issues raised by the consultation exercise concerning why mixed mode survey designs have, or have not, been used and consider what are the key advantages, obstacles and pressures surrounding the decision to mix modes.

Causes of Mode Effects: Separating out Interviewer and Stimulus Effects in Comparisons of Face-to-face and Telephone Surveys

Caroline Roberts; City University, United Kingdom

Annette Jäckle; University of Essex, United Kingdom

Peter Lynn; University of Essex, United Kingdom

This paper presents findings from an experimental study carried out in the context of the European Social Survey, to assess the impact a change in data collection mode from the current face-to-face interviewing to telephone might have on data quality and to study the likely causes of any observed mode effects. Evidence from previous studies suggests which differences in response can be expected between telephone and face-to-face interviewing and also suggests likely causes of such differences. Previous empirical studies are, however, often limited in their ability to isolate different causes. The experimental design enabled us to distinguish mode effects caused by differences in the type of question stimulus used in each mode (audio vs. visual) and mode effects caused by the presence or absence of the interviewer. The design included three comparison groups (two interviewed face-to-face (one with showcards, one without) and the third by telephone). We found evidence of effects caused by the presence of the interviewer, but few stimulus effects. We tested a number of hypotheses about the likely causes of mode effects on response, focusing on three forms of satisficing and social desirability bias. We found no evidence that using showcards influenced response quality, either positively or negatively. Unlike previous studies, we found no support for the hypothesis that telephone respondents were more likely to satisfice. However, consistent with our expectations, we did find telephone respondents were more likely to give socially desirable responses across a range of indicators.

Survey Errors in Mixed-mode Pilots That Incorporate Face-to-face, Telephone, Paper and Web Interviewing

Björn Janssen; Statistics Netherlands (CBS), Netherlands

Barry Schouten; Statistics Netherlands (CBS), Netherlands

Face-to-face, telephone and paper interviewing are well-established data collection modes in survey research. There is a vast literature on how to estimate population statistics using these modes. In the last decade, also web interviewing has become a regular mode for interviewing, and the literature on this mode is growing rapidly. The rise of web surveys also raised the question at many survey organisations whether the standard face-to-face and telephone surveys can be replaced by cheaper combinations of data collection modes while maintaining the same quality level of data.

Combining data collection modes, however, affects all survey errors, from measurement error to unit non-response. The question, therefore, translates to total survey design and is not a simple one to answer; it is almost unfeasible to fully disentangle the various errors that contribute to the total survey error while shifting to another data collection strategy.

Over the last three years, Statistics Netherlands conducted four pilots that involved both sequential and concurrent mixes of the modes face-to-face, telephone, paper, and web. The pilots were run parallel to existing household surveys. In the presentation, we will discuss the various components of the survey error for the chosen mixed-mode strategies. Furthermore, we will identify a number of research questions that need attention in the nearby future.

Measurement Equivalence of Paper-and-pencil and Online Organizational Surveys: a Large Scale Examination in 16 Countries

Alain De Beuckelaer; University of Gent, Belgium

Filip Lievens; University of Gent, Belgium

In multinational surveys, mixed-mode administration modes (i.e., combining online and by paper-and-pencil administration) are increasingly used. To date, no studies have investigated whether measurement equivalence exists between online data collection and data collection using the conventional paper-and-pencil method in organizational surveys which include a large number of countries. This paper examined the measurement equivalence of a truly global organizational survey across online and paper-and-pencil survey administrations. Data from an organizational survey in 16 countries (N=52,461) across the globe were used to assess the measurement equivalence of an organizational climate measure within each country in which the survey was administered. The empirical data provided strong indications

which support the measurement equivalence of the multi-item survey instrument across online and paper-and-pencil surveys in virtually all countries in which the survey was conducted. These findings suggest that merging data obtained through online and paper-and-pencil data administration in a particular country is legitimate as no evidence was found for differential effects across both modes of data collection.

7 Specific Topics I

Sessions in Specific Topics I

Emotional Intelligence Research	190
Demographic and Socioeconomic Variables in Cross-national Surveys	193
Cost-effective Data Editing for Business Surveys	198
The Impact of Media-reported Events on Attitudes and Opinions: Measurement and Analysis	201
Network Survey Measurement	207
Statistics on Income and Living Conditions (EU-SILC)	212
Income Measurement in Surveys	215
Survey Measurement of Service Quality	221

7.1 Emotional Intelligence Research

Joan Manuel Batista-Foguet; ESADE, Universitat Ramon Llull, Spain

Presentations

1. *Why ESCI Response Modality Should Be Stretched for Measuring Emotional Competencies?* – J. Batista-Foguet (Spain), L. Guillén (Spain), W. Saris (Spain), R. Serlivos (Spain), R. Boyatzis (United States)
2. *Construct Validity of a Behavioral Approach to Emotional Intelligence* – L. Guillén (Spain), W. Saris (Spain), J. Batista-Foguet (Spain), R. Serlivos (Spain), R. Boyatzis (United States)
3. *Enhancing Experiential Learning in Management Education* – R. Serlivos (Spain), A. Leaverton (Spain), L. Guillén (Spain), J. Batista-Foguet (Spain), R. Boyatzis (United States)

Presentations in the session “Emotional Intelligence Research”

Why ESCI Response Modality Should Be Stretched for Measuring Emotional Competencies?

Joan Manuel Batista-Foguet; ESADE, Universitat Ramon Llull, Spain

Laura Guillén; ESADE, Universitat Ramon Llull, Spain

Willem E. Saris; ESADE, Universitat Ramon Llull, Spain

Ricard Serlavos; ESADE, Universitat Ramon Llull, Spain

Richard Boyatzis; Case Western Reserve University, United States

Boyatzis and Goleman proposed the use of the Emotional Competency Inventory, a self-report and informants’ measure that assesses how the person expresses their handling of emotions in life and work settings (Boyatzis and Sala, 2004). The ECI-2 is constituted by manifest behaviors that are shown in daily work situations.

Boyatzis and Goleman have recently proposed the introduction of a reviewed version of the questionnaire, the Emotional and Social Competency Inventory (ESCI). The ESCI version of the questionnaire has 60 items with response categories based on frequency of demonstration or observation. Each item is on a scale of 1 to 5 and an optional answer of ‘I don’t know’ or ‘I have not had the opportunity to observe the person in an appropriate setting’ is read into the data as blank. The use of this scale leads to ordinal data for which some of the usual data analysis procedures are not justified attenuating most of the estimates. This paper stretches the response modality to eleven points scale and discusses the relative merits of it.

This study was conducted as part of a learning project at ESADE (Barcelona, Spain) based on the self-directed learning theory (Boyatzis, 2002; Boyatzis et al., 2000). Participants were MBA students of four different programs. A total of 400 participants rated themselves on the ESCI via a web-platform that permits random assignments to the two conditions/methods – five and eleven points scales. We designed a split-ballot MultiTrait-MultiMethod (SB-MTMM) experiment with four groups for the two methods and their replications.

The proposed eleven points scale has shown to provide better measurement quality regarding both Psychometric properties and Information Theory indices.

Construct Validity of a Behavioral Approach to Emotional Intelligence

Laura Guillén; ESADE, Universitat Ramon Llull, Spain

Willem E. Saris; ESADE, Universitat Ramon Llull, Spain

Joan Manuel Batista-Foguet; ESADE, Universitat Ramon Llull, Spain

Ricard Serlavos; ESADE, Universitat Ramon Llull, Spain

Richard Boyatzis; Case Western Reserve University, United States

The concept “emotional intelligence” (EI) resonates in the business world (Domagalski, 1999), and many authors have called for more research that clearly conceptualizes it (Sala, 2002; Becker, 2003, Day and Carroll, 2004). Some authors claimed that some EI models contain an incomprehensible mix of abilities, behaviors and general dispositions that merge personality attributes with mental ability (e.g. Mayer et al., 2000). Within the controversy of defining EI, the behavioral approach involving the definition and measurement of EI in terms of competences, has not received much attention. The rationale of studying the appropriateness of the emotional competency model seems relevant and will serve to increase understanding of the field as a whole and prove its usefulness within educational settings.

Boyatzis and Goleman proposed the use of the Emotional Competency Inventory that assesses how a person expresses their handling of emotions in life and work settings (Boyatzis and Sala, 2004). The most notable limitation of the ECI is the lack of factorial validity (e.g. Boyatzis and Sala, 2004,

Pérez et al., 2004). Boyatzis and Goleman proposed recently the introduction of a reviewed version of the questionnaire, the Emotional and Social Competency Inventory (ESCI). The test specifies twelve competencies which are grouped into four theoretical clusters, but Boyatzis, Goleman and Rhee (2000) claimed that competencies may be clustered empirically in more functional ways.

The objective of this paper is to study the underlying structure of the emotional and social competencies as conceptualized by Boyatzis and Goleman and to assess their validity and reliability. This involves discussing threats to construct validity detected so far in the literature and the proposal of alternative ways to overcome them. As suggested by Guillén (2007), it is hypothesized that the competencies are related to the social motives of power, achievement and affiliation which drive people to act in a particular way thus energizing human behavior (McClelland, 1980, 1985, 1987).

This paper is based on the others' scores of the ESCI of 600 MBA students at ESADE, Spain, that were involved in the LEAD course of personal and professional development based on emotional and social competencies as proposed by Goleman, Boyatzis and McKee (2002).

Enhancing Experiential Learning in Management Education

Ricard Serlavos; ESADE, Universitat Ramon Llull, Spain

Amy Leaverton; ESADE, Universitat Ramon Llull, Spain

Laura Guillén; ESADE, Universitat Ramon Llull, Spain

Joan Manuel Batista-Foguet; ESADE, Universitat Ramon Llull, Spain

Richard Boyatzis; Case Western Reserve University, United States

Research on experiential learning has gained new importance due to recent efforts to improve management education. D. Kolb's (1984) experiential learning theory (ELT) is one of the most influential theories applied to management development. According to ELT, learning is described as a continuous, cyclical process of responding to diverse personal and environmental demands that arise from the interaction between concrete experience, reflective observation, abstract conceptualization, and active experimentation (Kayes, 2002). The resulting 'learning style' (LS) describes individual differences in learning based on the learner's preference for employing the different phases of the learning cycle. Four learning styles have been identified: accommodating, diverging, assimilating and convergent.

Inspired by the innovative experience pioneered by CWRU in Cleveland, Ohio (Boyatzis, Cowen & Kolb, 1995), ESADE (Barcelona, Spain) is engaged in an institutional endeavour to promote experiential learning in its MBA programs by creating the appropriate environmental conditions.

The paper is based on the scores of the Learning Style Inventory (LSI) of a sample of 650 participants from four MBA (full and part time) and ten Executive Education different programs. The LSI is a self-report instrument developed by D.Kolb to assess individual learning styles (Kolb 1971, 1999) and designed for students and managers to assess their learning along the four dimensions of experiential learning. Past research with the LSI has identified learning style differences among different academic backgrounds.

Within the context of this project, the objective of this paper is to increase understanding of LS and its influence on management development programs as well as analyze the relationship between LS and some relevant variables. Specifically we will show how academic background is related with a specific LS and how the latter is related to participant's philosophical orientations, as defined and measured by Boyatzis's Philosophical Orientation Questionnaire (POQ).

7.2 Demographic and Socioeconomic Variables in Cross-national Surveys

Session coordinated by:

Christof Wolf; Centre for Survey Research and Methodology (GESIS-ZUMA), Germany

Presentations

1. *The Background of Background Variables in the International Social Survey Programme (ISSP)* – E. Scholz (Germany)
2. *Measuring Social Cohesion in Europe* – M. Mascherini (Italy), M. Nardo (Italy), D. Vidoni (Italy)
3. *The Relevance of 'private Household' for Cross-cultural Survey Comparison* – U. Warner (Luxembourg), J. Hoffmeyer-Zlotnik (Germany)
4. *How to Survey Education for Cross-national Comparisons: the Hoffmeyer-zlotnik / Warner Matrix of Education* – J. Hoffmeyer-Zlotnik (Germany), U. Warner (Luxembourg)
5. *The Dynamics of Mobility Capital: a Study of European Temporary Labour Mobility* – I. Harsløf (Norway), D. Zuev (Russia)
6. *Multiple Indicator Measurement of Occupational Status* – H. Ganzeboom (Netherlands), J. de Vries (Netherlands)

Presentations in the session "Demographic and Socioeconomic Variables in Cross-national Surveys"

The Background of Background Variables in the International Social Survey Programme (ISSP)

Evi Scholz; Centre for Survey Research and Methodology (GESIS-ZUMA), Germany

In cross-cultural surveys the comparability of information collected in different cultural contexts is crucial. The ISSP, as a result from its historical development, follows two different strategies of data collection: concerning the ISSP substantive variables input harmonisation is applied; concerning the ISSP background variables a special ex-post harmonisation approach is used. While the ISSP has fixed a set of obligatory background variables with stipulated measurement goals, question wording, question order, and the construction into the ISSP background variables is country-specific. Thus, having information on the harmonisation process is a major issue for ISSP researchers in order to reasonably decide on the comparability of the ISSP background variables. Since documentation on the construction of ISSP background variables is incomplete for most ISSP modules and countries and since concerns about implementation, compliance and completeness have grown, the ISSP has started several initiatives to improve its background variables.

This presentation starts with details of the harmonisation strategy of background variables in the International Social Survey Programme (ISSP). Then it reports on current developments, in particular the extra-ordinary and extensive documentation of background variables that the ISSP has introduced starting with its 2004 module. This unique documentation of the harmonisation allows researchers to follow the construction of the background variables from their very beginning. A special documentation template is introduced and explained; it was designed to cover all aspects that might be of interest for cross-national researchers using ISSP background variables. Components of the template are, for example, the original question wording or the bridging from country-specific to ISSP harmonised background variables. Finally, the first experiences from the new ISSP documentation will be presented.

Measuring Social Cohesion in Europe

Massimiliano Mascherini; European Commission - Joint Research Centre, Italy

Michela Nardo; European Commission - Joint Research Centre, Italy

Daniele Vidoni ; European Commission - Joint Research Centre, Italy

The Treaty Establishing the European Community sets Social Cohesion as one of the main EU objectives. Still, no agreement has yet been reached on the definition of the concept. The OECD, the council of Europe, and the European Union have adopted a common position by which "social cohesion is the capacity of a society to ensure the welfare of all its members, minimizing disparities and avoiding polarization. A cohesive society is a mutually supportive community of free individuals pursuing these common goals by democratic means" (Council of Europe, 2004). The World Bank, on the other hand, insists on the fact that "social cohesion is the ongoing process of developing a community shared values, shared challenges and equal opportunity. . . , based on a sense of trust, hope, and reciprocity. . ." (Social Cohesion Network, Policy Research Initiative).

While the Council of Europe looks mainly at the stock dimension of cohesion and points to its economic and social components, the World Bank highlights the dynamics behind its development and points to the aspects related to fostering network formation, community ties and institutions.

By using mainly 2002 data from Eurostat and the OECD on the majority of the EU27 Countries, the proposed project builds on the foresaid definitions and proposes a Composite Indicator of Social Cohesion (CISC) organized around a flux and a stock pillars.

The robustness of the Composite Indicator of Social Cohesion is tested in different ways by using Factor Analyses on the available data and by performing sensitivity analysis on a plurality of scenarios (all with their implications in terms of standardisation, weighting schema and alternative ways of composing the composite indicator).

Subsequently, in order better to understand the phenomenon of Social Cohesion, the CISC and its components are independently compared to other social and economic indicators such as GDP per capita, Corruption Perceptions Index, Human Development Index, and the Global Gender Gap Index.

The Relevance of 'private Household' for Cross-cultural Survey Comparison

Uwe Warner; Centre d'Études de Populations, de Pauvreté et de Politiques Socio-Economiques (CEPS / INSTEAD), Luxembourg

Jürgen H.P. Hoffmeyer-Zlotnik; Centre for Survey Research and Methodology (GESIS-ZUMA), Germany

Different cultures and nations use their relative definition of 'private household'. They correspond to the cultural and national structures of social life. These differences result in diverse household compositions and unequal sizes across European nations.

Comparing household measures over countries survey analysts face several inconveniences. The composition of the surveyed household has direct impact on the respondents answer about the household size. Therefore an effect on total net household income is visible.

Different membership rules constitute also differences in the socio economic status of the individual household members.

In a first step we summarize definitions of household used in national surveys across Europe. Same dwelling, sharing economic resources, common housekeeping and family ties are the main and mostly used criteria.

In a second step we discuss the possible combinations of these elements and the strategies of operationalization in social surveys.

The third part illustrates the findings. We use ESS, ECHP and administrative micro data from official statistics. The country differences become obvious.

Our conclusion is a revised fieldwork instrument measuring household in social surveys that increases data comparability across cultures and countries.

How to Survey Education for Cross-national Comparisons: the Hoffmeyer-zlotnik / Warner Matrix of Education

Jürgen H.P. Hoffmeyer-Zlotnik; Centre for Survey Research and Methodology (GESIS-ZUMA), Germany

Uwe Warner; Centre d'Études de Populations, de Pauvreté et de Politiques Socio-Economiques (CEPS / INSTEAD), Luxembourg

Social surveys collect information on socio-demographic characteristics of respondents eligible for the interview. Among others, the highest attained level of education is one of the variables explaining the respondent's transition from school to work, the position in the labour force and its segments, the social and economic behaviour of the individual actor and structural inequalities in modern societies. Manifold strategies to operationalize the qualification and education variables during interviews can be observed in social surveys. They differ in the underlying latent concept captured, the ranking and classifying of levels into categories and clusters, the degree of classification and measurement details and finally the capabilities of comparison across time and across nations and cultures.

The measurement of education for comparative research across countries is a complex task. The national systems of education and schooling are differently organized across the nations. Altogether four different types of school and training systems can be identified in Europe. In this paper we will sort the national certificates from general and vocational schools into one matrix, the newly developed Hoffmeyer-Zlotnik/Warner-matrix of education. This matrix allows us to compare the highest level of education a person has reached, as a combination of general and vocational education and usable for a person to obtain a starting position on the labour market. This article discusses those measurement instruments normally used in international comparative surveys and introduces the Hoffmeyer-Zlotnik/Warner-matrix of education, contrasting this matrix with the other established measurement

instruments. To demonstrate the validity of our matrix, we show the advantages of our matrix exemplified in one case from nations out of the four types of different educational systems: Germany, Luxembourg, Denmark, and France.

The Dynamics of Mobility Capital: a Study of European Temporary Labour Mobility

Ivan Harsløf; Oslo University College, Norway

Dennis Zuev; Krasnoyarsk State University, Russia

This paper addresses the issue of international temporary labour mobility as a potential resource for individual workers. The working hypothesis of this study is that employees' 'mobility power' is enhanced by international economic market integration. International work experience is expected to form convertible capital beneficial to the employee on return to her country of origin. The purpose of the study is to identify and discuss European differences in international mobility patterns and to examine the impact of international work experience on people's current working and living conditions. Through analyses of European Social Survey data (2nd round), an assessment of the manner in which temporary, international work experience affects critical aspects of people's life including 'employability', quality of current job, social network (social capital), and family life, is presented.

The analyses suggest that there are neither positive nor negative effects of having international work experience on one's current employment chances. Among people who are presently in employment, having work experience from abroad is associated with decreased odds of being member of a trade union. In terms of social network, the internationally mobile group is found to be more active in socialising with relatives, friends, and colleagues than are their stay-at-home counterparts. Singles are overrepresented among internationally mobile workers. In summary, while associated with stronger networking activities, temporary work abroad seems to indicate a more individualised way of life, abandoning binding communities as represented by trade unions and intimate relationships. In a discussion of the interrelation between micro and macro level social capital, such a finding should be of concern.

Multiple Indicator Measurement of Occupational Status

Harry B.G. Ganzeboom; Vrije Universiteit Amsterdam, Netherlands

Jannes de Vries; University of Tilburg, Netherlands

Occupational status is most often measured using single indicators. These indicators are implicitly assumed to have sufficient reliability and validity, despite reports in the literature that there are many imperfections in often used measures. Such results are almost always obtained using multiple sources or 'proxy' measurement. In this paper we examine a methodology to use parallel measures in a single survey from the same source and estimate the reliability and validity of occupational status using a MTMM (Multi-Trait Multi-Method) model. The parallel measures are generated by asking about occupation in a crude (pre-coded) format and in a detailed (verbatim, post-coded) format. The paper builds upon work reported by Ganzeboom (2005) on a similar design in the ISSP 1987, but uses data collected in five recent surveys in the Netherlands that include parallel measures for fathers's and mother's occupation, and respondent's first and current occupation.

We find very little evidence of systematic measurement error in either crude or detailed measures. This is so for methods effects (crude versus detailed), but also for education bias (a form of correlated error). Like in the previous analysis by Ganzeboom (2005), we find that crude and detailed measure are subject by about the same level of random error. Despite observed reliability meeting often-quoted benchmarks such as $\alpha > 0.70$, we also find that correction for attenuation brought about by random measurement error changes conclusions about the structure of status attainment substantially. In particular, our analysis replicates the finding in Ganzeboom (2005) that there are no direct effects education on earnings, once occupation is adequately controlled, i.e. by using multiple measurement.

References

Ganzeboom, Harry B.G. (2005). "On the Cost of Being Crude: A Comparison of Detailed and Coarse Occupational Coding." Pp. 241-258 in: Jürgen H.P. Hoffmeyer-Zlotnik, *Methodological Aspects of Cross-National Research*, Mannheim: ZUMA-Nachrichten [Special Issue #11], 2005.

7.3 Cost-effective Data Editing for Business Surveys

Session coordinated by:

Gary Brown; Office for National Statistics, United Kingdom

Presentations

1. *Designing the Editing and Imputation Process of Business Surveys: the Edimbus-manual* – B. Hulliger (Switzerland)
2. *Data Editing for a Large-scale Business Survey: a Comprehensive Review and Some Lessons Learned* – P. do Nascimento Silva (United Kingdom), A. Al-Hamad (United Kingdom), G. Brown (United Kingdom)
3. *A Quality-driven Approach to Reducing Data Editing Costs* – G. Brown (United Kingdom), A. Al-Hamad (United Kingdom), B. Martín (United Kingdom)

Presentations in the session “Cost-effective Data Editing for Business Surveys”

Designing the Editing and Imputation Process of Business Surveys: the Edimbus-manual

Beat Hulliger; University of Applied Sciences Northwestern Switzerland (FHNW), Switzerland

Raw data of business surveys often contains errors and missing values. Their detection and treatment is expensive and long. Many methods and approaches for particular problems exist but few comprehensive manuals are available. The ED-IMBUS project, a joint effort of the National Statistical Institutes of Italy (ISTAT), Netherlands (CBS) and Switzerland (SFSO) with support from Eurostat, develops a Recommended Practices Manual (RPM) which treats the design of the process of editing and imputation and its place in the design of a survey. The main objectives of editing and imputation, to inform on the quality of a survey and to prepare the data for analysis, are explained. Advantages and limitations of established methods for the detection and treatment of errors and missing values are discussed and recommendations on their use are given. The methods range from computer aided interactive treatment to automatic procedures and from micro to macro level. Testing of the editing and imputation process and monitoring its application by quality indicators as well as the documentation of editing and imputation are important parts of the EDIMBUS-RPM.

The talk will give an overview over the EDIMBUS manual and its main features. Selective editing, which aims at concentrating efforts on the most influential errors, and outlier detection and treatment are discussed in more detail. The relation between the score functions of selective editing, the influence of observations and outliers is explored.

Data Editing for a Large-scale Business Survey: a Comprehensive Review and Some Lessons Learned

Pedro Luis do Nascimento Silva; University of Southampton, United Kingdom

Alaa Al-Hamad; Office for National Statistics, United Kingdom

Gary Brown; Office for National Statistics, United Kingdom

Initial results from an ongoing project to review and improve the editing process adopted in a large-scale annual business survey are reported. The approach adopted for the review was holistic, aimed to long term future improvement of the survey process, and not focussed on short term cost savings. It comprised examining detailed specifications for the current editing process, and their connections to data collection instruments and data capture technology. Indicators of the impact of editing were obtained from sets of pre- and post-edited data, and related to several quality issues raised regarding the questionnaire design, instructions, and data capture. The analysis of these results suggests that the survey could benefit from a combination of questionnaire and data collection redesign, as well as from improved editing and imputation methods. If editing is to lead to quality improvements then the editing process must be redesigned to produce two equally important sets of outcomes: the usual ‘clean’ or ‘fit for purpose’ datasets, and an ‘analytic report’ summarising the editing performed and its impact on the data, and connecting these to the relevant aspects of the survey design. This analytic report is the key to achieving long term cost-efficiency of survey editing, because it would provide crucial information needed to reduce or avoid the same types of errors appearing in the survey data in the first place.

A Quality-driven Approach to Reducing Data Editing Costs

Gary Brown; Office for National Statistics, United Kingdom

Alaa Al-Hamad; Office for National Statistics, United Kingdom

Begoña Marfín; Office for National Statistics, United Kingdom

The Office for National Statistics, UK, has developed a new methodological approach to achieving data editing efficiency in business surveys by linking reductions in editing costs to the impact on quality. Known as SNOWDON, this approach is already somewhat a victim of its own success. This success has resulted in its long-term development temporarily taking a back seat to short-term implementation, as survey managers need to make savings now. Given that SNOWDON delivers savings with the minimum impact on survey results, it has been appropriate to direct methodological resources towards achieving these targets in the short-term.

However, even though SNOWDON is currently in use, it has been possible to continue some practical development – extending its range to include optimisation of selective editing thresholds in addition to optimal re-parameterisation of editing rules.

This talk explains how SNOWDON works, shares some results, and outlines plans to develop it into an interactive tool that will allow survey managers to optimise parameters and thresholds themselves.

7.4 The Impact of Media-reported Events on Attitudes and Opinions: Measurement and Analysis

Session coordinated by:

Ineke Stoop; Social and Cultural Planning office (SCP), Netherlands

Presentations

1. *Event Data and Social Surveys* – I. Stoop (Netherlands)
2. *Automatic Event Reporting: Some Results of a Small Project* – C. Züll (Germany), J. Landmann (Germany)
3. *Extending Event and Claims-making Analysis As a Tool for Cross-national Survey Research* – P. Statham (United Kingdom)
4. *Early Respondents, Late Respondents and Traumatic Events in the Midst of Fieldwork* – J. Verhagen (Netherlands)
5. *Event Data: Assessing the Impact of Headline News in Values, Attitudes and Opinions.* – A. Torres (Portugal), R. Brites (Portugal), I. Cardoso (Portugal)
6. *The Identification of Immigration Issues Through the Use of Event Data.* – T. Stathopoulou (Greece)
7. *Natural Disaster and Its Reflections in Opinions and Attitudes of ESS 2 Respondents in Slovakia* – J. Výrost (Slovakia), M. Kentoš (Slovakia), D. Fedáková (Slovakia)
8. *Event Data: Assessing the Impact of Headline News in Values, Attitudes and Opinions.* – A. Torres (Portugal), R. Brites (Portugal), I. Cardoso (Portugal)

Presentations in the session "The Impact of Media-reported Events on Attitudes and Opinions: Measurement and Analysis"

Event Data and Social Surveys

Ineke Stoop; Social and Cultural Planning office (SCP), Netherlands

The death of a loved one, the loss of a job or being victim of a serious crime may have a large impact on individual well-being and attitudes. The contextual impact on individual response behaviour will not create major effects, however, as long as the contexts and events vary individually in an idiosyncratic fashion. The impact of events must be considered and, whenever possible, controlled as soon as whole societies are thus influenced in a way, which is not uniform across countries. This is especially true in general, longitudinal studies.

For this reason, from the beginning of the study, an inventory of major, national events that happened during fieldwork was foreseen in the European Social Survey. Event reporting in the ESS started in Round 1 (2002/2003) with rather free-format guidelines asking National Coordinators and their teams in each participating country to send in monthly reports on events that received 'prominent attention' in national newspapers. This was defined to mean 'front page news' or 'appearing regularly in larger articles on later pages' on several days. The selection of newspapers was more or less left to the national reporter. Reporters assigned events to fixed categories, provided keywords and a description, gave a start date and end date (if possible), mentioned the source, and assessed the likely size and direction of the event's impact on the survey answers.

In Round 2 (2004/2005) the guidelines were revised resulting in a more standardised format. Events were reported weekly, rather than monthly. As in Round 1 incoming event reports were posted on a website (accessible via <http://ess.nsd.uib.no>) to give an up-to-date overview of incoming event reports, and also to give access to guidelines, information notes and background information. This transparency was helpful not only to users who wished to have an overview of weekly or monthly events in each country, but also to reporters themselves as a way of checking how their colleagues in other countries were using the system.

In Round 3 (2006/2007) ESS-eventnet was implemented (www.scp.nl/ess/eventnet), a web tool that allows national event reporters to upload events themselves. The reported events are now part of a database. ESS-eventnet provides access to reports on each individual event and can also generate a number of simple reports. Practical improvements are foreseen for Round 4 (2008/2009).

The presentation will give an overview of event reporting in the European Social Survey, give an impression of the wealth of data that is available, discuss how event reporting can be improved and discuss possibilities and problems when relating events to survey data

Automatic Event Reporting: Some Results of a Small Project

Cornelia Züll; Centre for Survey Research and Methodology (GESIS-ZUMA), Germany

Juliane Landmann; Centre for Survey Research and Methodology (GESIS-ZUMA), Germany

Societal events like elections, significant changes in laws, demonstrations, but also extreme weather conditions can have an effect on a society and, consequently, influence the attitudes of its population. The assumption for surveys is that respondent behaviour or answers to some questions are influenced by significant societal events in various areas. When conducting a survey, the impact of an event must be considered and, whenever possible, controlled, especially in survey projects in which different countries participate. As part of the first three rounds of the ESS project, National Coordinators have collected information on major events which took place during the fieldwork. However, manual identification of significant events which occur during the data collection phase is an error-prone and very time-consuming task. Therefore, we have developed a procedure to identify events using computer-assisted content analysis. We tested this procedure to determine its usability in a small project. The procedure is based on a combination of two techniques of computer-assisted text analysis: a reference

text technique and the statistical association approach. First, we will introduce our procedure based on a reference text corpus and word co-occurrences. Secondly, we will present some of the results of identifying events automatically in Germany as well as in Great Britain. As text basis for identifying events, we use newspaper articles published in September to November 2006 at the beginning of round 3 of the ESS. Finally, we will discuss the advantages and risks of such an automatic procedure as well as the possibility of combining such an application with further (manual) coding.

Extending Event and Claims-making Analysis As a Tool for Cross-national Survey Research

Paul Statham; University of Bristol, United Kingdom

Claims-making analysis is an established approach for examining the public dimension of politics (Koopmans and Statham 1999, Koopmans, Statham et al 2005) from newspaper sources. Following its 'protest event' analysis roots in social movement research, the approach takes news as a 'source' for claims-making by reported 'third party' actors, with journalists' own opinions and comments on actors and events being bracketed out. Basically, news is taken as a source recording political events and claims in the public domain. For claims which are retrieved from reports in the main news section, the unit of analysis is a claims-making act and not an article. A claims-making act (shorthand: a claim) is a strategic action in the public sphere. It consists of intentional public speech acts which articulate political demands, calls to action, proposals, and criticisms, which, actually or potentially, affect the interests or integrity of claimants and/or other collective actors in a specific issue-field. The method codes a wide range of actors including civil society groups, such as employers, trade unions, and NGOs, and state actors, such as courts, legislatures, governments and supranational institutions. The approach has been successfully applied in several cross-national comparative research projects. The challenge now is to examine to what extent it is suitable for controlling for 'exceptional' events that may shape or bias responses to surveys at specific times within a cross-national framework. This requires bringing the needs of cross-national survey research together with the efficacy of the existing method thereby leading to an extension and development.

Early Respondents, Late Respondents and Traumatic Events in the Midst of Fieldwork

Josine Verhagen; Social and Cultural Planning office (SCP), Netherlands

Events that happen in the fieldwork period of a survey can have a profound effect on survey variables. During the fieldwork period of the Cultural Changes survey 2004 in the Netherlands, the famous filmmaker Theo van Gogh was murdered by a muslim fundamentalist because of his extreme statements about muslims and the islam. In this case it can be shown that the murder of Theo van Gogh influenced the importance people attached to the protection of freedom of speech as a political goal. The mean ranking of the protection of freedom of speech as a political goal increased directly after the murder and then gradually decreased after approximately four weeks, suggesting that interviewing people during that period would give an inflated estimate of the real importance attached to the protection of freedom of speech by Dutch citizens.

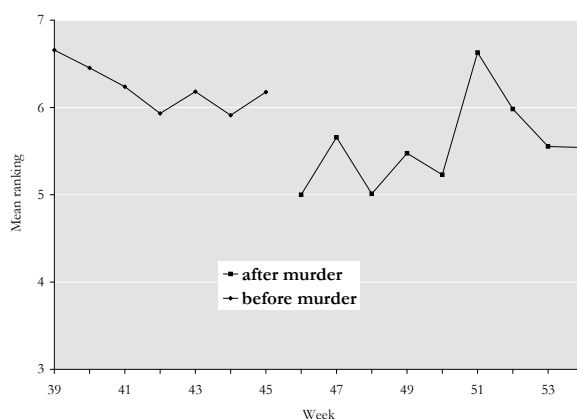
The design of this small natural experiment was confounded, however. The persons interviewed before the murder were easy to catch at home and willing to participate in the survey on a short notice. The persons who were interviewed after the murder were mostly persons who needed multiple visits before they were found at home and cooperated. Research in this area (Stoop, 2005; Verhagen, 2007) showed that those hard-to-reach respondents are more likely to be younger, to be higher educated and to have an active pattern of activities that take one outside the home. Although Verhagen (2007) found that this didn't influence the survey variables much, it has to be taken into account that different people with possibly different opinions were interviewed after the murder, and that this could be part of the explanation of the enhanced attached importance to the freedom of speech.

An attempt will be made to disentangle the effects of the murder and contactability. Both have strong effects on the attached importance to the freedom of speech, which decrease in strength considerably

when corrected for the other. The effect of the murder remains significant, however. Age and education do not influence the murder effect, while these variables do explain the effect of contactability. It can be concluded that the effect of the murder is real and that the effect of contactability is explained partly by this murder effect and partly by age and education.

Although in this case the effect of the van Gogh murder could not be explained by confounding variables, it is advised to keep in mind that event studies can be confounded by the field work set-up if the persons interviewed before the event differ from the persons interviewed after the event. This difference can be caused by non-response factors like contactability or increased efforts to include refusals after a certain period, but also by other factors like fieldwork that is unevenly spread across areas.

On the other hand research into non-response has to take into account the effects of events. A wrong conclusion could have been reached here: that respondents who are difficult to reach attach more importance to the freedom of speech as a political goal.



References

- Stoop, Ineke A.L. (2005) *The Hunt for the Last Respondent*. The Hague: SCP.
 Verhagen, A.J.(2007) *Robuuste meningen? Het effect van responsverhogende strategieën bij het onderzoek Culturele Veranderingen in Nederland*. The Hague: SCP.

Event Data: Assessing the Impact of Headline News in Values, Attitudes and Opinions.

Analia Torres; Instituto Superior de Ciências do Trabalho e da Empresa - ISCTE, Portugal

Rui Brites; Instituto Superior de Ciências do Trabalho e da Empresa - ISCTE, Portugal

Inês Cardoso; Instituto Superior de Ciências do Trabalho e da Empresa - ISCTE, Portugal

In this paper we will analyse the role of Event Data Tool in accessing European's attitudes and opinions. This useful tool - which consists in the compilation of newspapers headlines in the various countries, during European Social Survey (ESS) fieldwork - aims to assess the impact of headline news in the answers given by individuals to social and political issues. Does the announce of a political fraud in a certain country, for instance, lead to very negative opinions about that specific national government's performance?

Measuring the effect of that news is a very complex task since, first, there are several factors shaping people's views and secondly because the questions for capturing values, attitudes and opinions in ESS are of a very wide range so being very difficult to assert the impact of event news.

We can't state, though, that person's beliefs were undoubtedly induced by media, but we can examine European's expressed trust in national and international institutions and also detect changes in those opinions, trough the comparison of 2002 and 2004 (rounds 1 and 2 of the ESS) results. At the

national level, we will try to perceive, namely, the impact of the government dismissal on the Portuguese responses. From the analysis of the main events that happened during the fieldwork period, we can also examine Europeans' values and attitudes for other themes like homosexuality, maternity, legal system and democracy. We also intend to establish some hypothesis of what expect to find in 2006 results, considering the main events that have occurred last year.

The Identification of Immigration Issues Through the Use of Event Data.

Theoni Stathopoulou; EKKE, Greece

The basic aim of this analysis is to explore the extent to which immigration and minority issues concern EU countries, that participated in the European Social Survey, through the utilisation of event data (Rounds 2 and 3). A fundamental premise for our analysis is that the degree to which such events become important is illustrated by the degree to which the Mass Media present them. More specifically, this analysis focuses on the following questions:

1. To what extent do participating countries address immigration and/or minority issues?
2. Do some countries regularly address issues of immigration policy and/or minorities more often than others?
3. Is it possible that participating countries address immigration policy issues only as a result of the occurrence of important immigration events?
4. Do immigration issues within one country also affect neighbouring countries?
5. What were the most important immigration and/or minority issues that concerned EU participating countries for the period covered by Rounds 2 and 3 of ESS event data collection? How long did they last? In which countries did events occur, which countries appear to have been affected and how soon after the event?

We will attempt to provide answers to the above questions through the use of appropriate methods of textual analysis.

Natural Disaster and Its Reflections in Opinions and Attitudes of ESS 2 Respondents in Slovakia

Jozef Výrost; Slovak Academy of Sciences, Slovakia

Michal Kentoš; Slovak Academy of Sciences, Slovakia

Denisa Fedáková; Slovak Academy of Sciences, Slovakia

During the fieldwork period of ESS 2 in Slovakia the wind storm in the most attractive Slovak mountains destination The High Tatras occurred (on 19 November 2004). It lasted just a few hours but damaged more than 12,7 million m³ of forest and extent of damage is estimated at 1,8 milliards SKK.

This event was of a great significance for Slovak citizens. Media of all kinds had monitored the situation in the Tatras for at least one month. The reflections of the event in the responses of Slovak respondents are presented in the paper. On the three groups of respondents (interviews taken before the disaster, immediately after and some time after it) the impact of the event at the media watching, reading, political engagement and well-being was analysed.

Event Data: Assessing the Impact of Headline News in Values, Attitudes and Opinions.

Analia Torres; Instituto Superior de Ciências do Trabalho e da Empresa - ISCTE, Portugal

Rui Brites; Instituto Superior de Ciências do Trabalho e da Empresa - ISCTE, Portugal

Inês Cardoso; Instituto Superior de Ciências do Trabalho e da Empresa - ISCTE, Portugal

In this paper we will analyse the role of Event Data Tool in accessing European's attitudes and opinions. This useful tool - which consists in the compilation of newspapers headlines in the various countries, during European Social Survey (ESS) fieldwork - aims to assess the impact of headline news in the answers given by individuals to social and political issues. Does the announce of a political fraud in a certain country, for instance, lead to very negative opinions about that specific national government's performance?

Measuring the effect of that news is a very complex task since, first, there are several factors shaping people's views and secondly because the questions for capturing values, attitudes and opinions in ESS are of a very wide range so being very difficult to assert the impact of event news.

We can't state, though, that person's beliefs were undoubtedly induced by media, but we can examine European's expressed trust in national and international institutions and also detect changes in those opinions, trough the comparison of 2002 and 2004 (rounds 1 and 2 of the ESS) results. At the national level, we will try to perceive, namely, the impact of the government dismissal on the Portuguese responses. From the analysis of the main events that happened during the fieldwork period, we can also examine Europeans' values and attitudes for other themes like homosexuality, maternity, legal system and democracy. We also intend to establish some hypothesis of what expect to find in 2006 results, considering the main events that have occurred last year.

7.5 Network Survey Measurement

Session coordinated by:

Anuška Ferligoj; University of Ljubljana, Slovenia

Presentations

1. *Quality of Dyadic Relations and Proxy Nonresponse in an Egocentric Network Study-dyadic Relations and Proxy Nonresponse in an Egocentric Network Study* – V. Stocké (Germany)
2. *Measuring Ego-centered Social Networks: Do Cheaper and Low Respondent Burden Methods Provide Accurate Estimates of Network Composition?* – V. Hlebec (Slovenia), T. Kogovšek (Slovenia)
3. *The Effects of Social Networks on the Performance of Phd Students: the Case of Slovenia* – P. Zihel (Slovenia), H. Igljč (Slovenia), A. Ferligoj (Slovenia)
4. *Networks of Phd Students and Academic Performance: a Comparison Between Slovenia, Germany and Spain.* – A. Capó (Spain), A. Ferligoj (Slovenia), D. Krebs (Germany), J. Hoffmeyer-Zlotnik (Germany), T. Kogovšek (Slovenia), V. Hlebec (Slovenia)
5. *Duocentered Networks* – L. Coromina Soler (Spain), G. Coenders (Spain), A. Ferligoj (Slovenia), J. Guia (Spain)
6. *The Effect of the Relationship Between Phd Students and Their Mentors on Student's Performance* – U. Matelič (Slovenia), A. Ferligoj (Slovenia), F. Mali (Slovenia)

Presentations in the session "Network Survey Measurement"

Quality of Dyadic Relations and Proxy Nonresponse in an Egocentric Network Study

Volker Stocké; University of Mannheim, Germany

Egocentric network studies often rely on proxy-reports about the characteristics of the network's first order zone. These reports are, however, subject to a considerable degree of item nonresponse. The present paper tests within a multilevel framework for the interviewer-, respondent- and dyadic-level determinants for German primary school parents' failure to provide information about attitudes and beliefs of their egocentric social network. The total nonresponse was found to be due to very few refusals, but the respondents were substantially susceptible to answer 'don't know'. Thus, the analysed questions must be regarded as difficult but not sensitive. According to random coefficient models, all three levels significantly contributed to the variance in 'don't know' responses. However, differences between dyads explained most and interviewer differences the least amount of variability. None of the analysed interviewer fixed effects and only the respondents' age and social class proved to explain the prevalence of nonresponse. However, the quality of the dyadic relationship between ego and alter had a strong effect on the susceptibility to item nonresponse. We found the length and emotional closeness of the dyadic relationships on the one hand, and the frequency and multiplexity of contacts on the other, to form distinct dimensions. Both dimensions had a significant negative net effect on the prevalence of 'don't know' responses. The effect of these relational dimensions proved to be non-additive. A significant interaction effect indicated that, with increasing opportunities for information exchange, the closeness of the relationship was much less relevant for nonresponse. The consequences for the quality of egocentric network data are discussed.

Measuring Ego-centered Social Networks: Do Cheaper and Low Respondent Burden Methods Provide Accurate Estimates of Network Composition?

Valentina Hlebec; University of Ljubljana, Slovenia

Tina Kogovšek; University of Ljubljana, Slovenia

In measuring ego-centered social networks two general approaches can be distinguished. A very simple way to evaluate membership in social network is to ask an ordinary survey question where response categories are types relationships (e.g. partner, parents, children, friends, etc.). This approach is very appealing as it saves time and money. However, information obtained by this approach is very limited.

Most often, when evaluating ego-centered networks, name generator approach is used. The list of egos (respondents) is obtained in the first step. In the second step, existing ties are identified - all alters with whom the focal ego has some sort of relationship. When all ties are identified, the contents of ties and the characteristics of ties are assessed. In most cases the characteristics of the alters are also measured. The name generator approach yields more data and it is also of higher quality. However, it is very time and money consuming and it requires considerable effort from respondent, when it is applied in self-administered mode or complex coordination between interviewer and respondent, when it is applied in personal interviews (e.g. Kogovšek et al., 2002).

In series of studies network composition was estimated using both approaches. Test-retest and split-ballot experiments on convenience samples of respondents were used to assess stability of network composition. Findings are discussed with regard to survey complexity, respondent burden, costs and quality of network composition estimates.

The Effects of Social Networks on the Performance of Phd Students: the Case of Slovenia

Petra Zihlerl; CATI d.o.o., Slovenia

Hajdeja Igljič; University of Ljubljana, Slovenia

Anuška Ferligoj; University of Ljubljana, Slovenia

The aim of the talk is to examine the contribution of individual and contextual level variables on the performance of PhD students, who are actively involved in research groups at the universities or research institutes during their study. We are particularly interested in what kind of research contexts promote the performance of PhD students and how important are their effects compared to individual characteristics.

Among the individual level variables we look at the motivation of PhD students (the centrality of work in PhD students' life) and the academic performance of their supervisor. Among the contextual variables we look at various network characteristics of research groups such as size and heterogeneity of the group, strength of ties between students and other researchers, density of research groups, and the existence of structural holes.

We use the data from the Slovenian study of academic research groups conducted in 2003/2004. The sample of research groups is representative of the Slovenian research groups, which include Ph.D. students and their supervisors under the "junior researchers" program financed by the Slovenian Ministry of Technology and Higher Education.

The results of the analysis indicate that the size and heterogeneity of research groups and the number of structural holes in the research network of PhD students have strong impact on the students' performance. The relationship between the strength of ties and performance shows some signs of non-linearity: it is the moderate strength of social ties, which comes along the heterogeneity of contacts that is beneficial for the performance of Ph.D. students. Besides, network variables turn out to be as important as individual level variables.

Networks of PhD Students and Academic Performance: a Comparison Between Slovenia, Germany and Spain.

Aina Maria Capó; University of Girona, Spain

Anuška Ferligoj; University of Ljubljana, Slovenia

Dagmar Krebs; University of Giessen, Germany

Jürgen H.P. Hoffmeyer-Zlotnik; Centre for Survey Research and Methodology (GESIS-ZUMA), Germany

Tina Kogovšek; University of Ljubljana, Slovenia

Valentina Hlebec; University of Ljubljana, Slovenia

In this presentation we compare regression models obtained to predict PhD students' academic performance in universities of Slovenia, Spain and Germany. Explanatory variables are characteristics of the PhD students' research group understood as an egocentered social network, background and attitudinal characteristics of the PhD students and some characteristics of the supervisors. Academic performance was measured by the weighted number of publications.

A questionnaire was designed for PhD students. Most of the variables were easily comparable across universities due to the careful translation procedure and pre-tests. When direct comparison was not possible we created comparable indicators.

We used a regression model in which the country was introduced as a set of dummy coded variables including all possible interaction effects. The optimal transformations of the main and interaction variables are discussed.

Some differences between Slovenian, German and Spanish universities emerge while the effects of some variables on performance are quite stable. We plan a qualitative follow up to understand the reasons for the differences.

Duocentered Networks

Lluís Coromina Soler; ESADE, Universitat Ramon Llull, Spain

Germà Coenders; University of Girona, Spain

Anuška Ferligoj; University of Ljubljana, Slovenia

Jaume Guia; University of Girona, Spain

If a pair of individuals is central, to a research problem (e.g. husband and wife, PhD student and supervisor) “duocentered” networks can be defined as an extension of egocentered networks. This structure is based on two central egos instead of one in the egocentered network or n egos in the complete network. The key point for this kind of network is that ties exist between the two egos and between them and all alters, but the ties among alters are not considered. This type of network can also be considered as a compromise. Complete network measurements are often costly to obtain and tend to be of low quality (especially for peripheral actors or if proxies are used) and usually contain a large proportion of missing data. Egocentered network data are less costly and of a higher quality but a lot of information is lost, as this type of network assumes that only one individual is central in the study.

From the definition of duocentered networks, we develop new social network measures, some of which are based on the measures for complete networks such as degree, closeness centrality or density, while others are tailor-made measures for solving specific problems when using duocentered network structures.

The proposed measures are used for the analysis of the networks of Slovenian PhD students and their supervisors. We specify three regression models to predict PhD student’s academic performance based on these social network measures for different relations such as advice, collaboration, emotional support, and trust. The results show that the model with absolute duocentered network measures predicts performance best.

Finally, we specified three regression models with absolute measures for egocentered, duocentered and both structures. We found that duocentered network variables alone lead to a higher explained variance than using egocentered network measures.

The Effect of the Relationship Between Phd Students and Their Mentors on Student’s Performance

Uroš Matelič; CATI d.o.o., Slovenia

Anuška Ferligoj; University of Ljubljana, Slovenia

Franc Mali; University of Ljubljana, Slovenia

This research is a part of the INSOC project (International Network on Social Capital and Performance), which includes Belgium, Germany, Slovenia, and Spain. The aim of the INSOC project is to examine the role of the research groups’ social capital in explaining the performance of PhD students in different countries (Waege et al., 2000).

The paper is based on the theory of the creative knowledge environments (e.g., Elgar, Hemlin, Allwood & Martin, 2004). An empirical study was performed to study the effect of the relationship between the PhD student and his/her mentor, together with mentor characteristics, on the PhD student’s performance in Slovenia. The sample consisted of doctoral students enrolled in the third year of a graduate program in the academic years 1999/2000 and 2000/2001 who were financed by the Ministry of Higher Education and Technology under the auspices of the program called “Junior Researchers”. The list of the PhD students and their mentors was obtained from the Ministry. The boundaries of the research groups in which the PhD students worked were defined by their mentors in September and October, 2003. We were able to contact 204 mentors (the response rate was 86%). Among them, 192 were willing to define their research group and provided e-mail addresses of all members of the group. We contacted these researchers and asked them to participate in a web survey between January and April, 2004. The response rate was 61 % for the PhD students and 52% for the other members of the research groups.

In this paper, the PhD student–mentor diads are considered. One-hundred-seventeen PhD students and 102 mentors answered the questionnaire. At the end of the data collection phase, we obtained 60 complete diads.

The findings obtained by multiple regression analysis are that the PhD student's performance is better if his/her mentor gives more freedom to the student in their research work, if the mentor does not impose his/her ideas too much on the student, if the mentor gives advices when the student needs them, and if mentor's scientific performance is high.

7.6 Statistics on Income and Living Conditions (EU-SILC)

Session coordinated by:

Jean-Marc Museaux ; European Communities Statistical Office (Eurostat), Luxembourg

Presentations

1. *Eu-silc in Slovenia* – R. Inglic (Slovenia)
2. *Comparison of EU-SILC and Hbs with Particular Attention to Income Variables* – A. Rybkowska (Poland), D. Raczka-Vargas (Poland), A. Szukielojc-Bienkunska (Poland)
3. *Effects of Rotational Design and Attrition in Czech EU-SILC* – M. Mysíková (Czech Republic), T. Smetanová (Czech Republic), S. Tourek (Czech Republic), M. Zelený (Czech Republic)
4. *The Weighting Procedures of the Swiss Household Panel from 1999 to 2005, Study of the Attrition, Applications and Recommendations for Silc-switzerland* – E. Graf (Switzerland)

Presentations in the session "Statistics on Income and Living Conditions (EU-SILC)"

Eu-silc in Slovenia

Rihard Tomaz Inglic; Statistics Slovenia, Slovenia

The paper will present conducting of EU-SILC in Slovenia in 2005. EU-SILC is the first sample survey where we use different sources to compose the database from different sources. One source of the data is a classical survey with the questionnaires, but the majority of the data we extracted for sampling persons from registers and other administrative sources. Some administrative sources we have in the Statistical Office, some of them we get from other governmental institutions.

In the paper we will describe advantages and disadvantages of the use of register and administrative sources in the view of the quality of the data and timelines as well.

Comparison of EU-SILC and Hbs with Particular Attention to Income Variables

Anna Rybkowska; Statistics Poland, Poland

Danuta Raczka-Vargas; Statistics Poland, Poland

Anna Szukielojc-Bienkunska; Statistics Poland, Poland

EU-SILC was carried out for the first time in Poland in 2005. The reference period for income variables was 2004. Up to 2004 it was the household budget survey that provided the main source of income data being the source of data used for Laeken indicators measurement. The presentation will consist of three parts. The first part will concern the methodological solutions adopted for EU-SILC and HBS with the focus on the methodological differences regarding such aspects as the reference periods for income variables, data collection, imputation of missing data or selection method. The second part will present the incomes of households and household members by kind of income gained in 2004 and collected by both surveys. The third part will cover Laeken indicators measuring poverty based on the two surveys. Both in the case of income variables and Laeken indicators differences between EU-SILC and HBS data are indicated.

Effects of Rotational Design and Attrition in Czech EU-SILC

Martina Mysíková; Czech Statistical Office, Czech Republic

Tamara Smetanová; Czech Statistical Office, Czech Republic

Stepán Tourek; Czech Statistical Office, Czech Republic

Martin Zelený; University of Economics, Prague, Czech Republic

The design of EU-SILC in Czech Republic follows the integrated cross-sectional and longitudinal design using the four-year rotational panel scheme. The contribution will first focus on the implementation of this design and the effects it has on cross-sectional survey data. Panel attrition is then addressed as a specific problem, which could have some impact on the survey results. We investigate the first two waves of the EU-SILC survey in the Czech Republic and try to reveal the effect of non-response in the second wave. The first part will, based on the first year data, describe the characteristics of households, which refused the interview in the second wave. The final part will be focused on the assessment of impact of attrition on survey results and on year-to-year trends, which can be observed in the annual cross-sectional data.

The Weighting Procedures of the Swiss Household Panel from 1999 to 2005, Study of the Attrition, Applications and Recommendations for Silc-switzerland

Eric Graf ; Swiss Household Panel, Switzerland

The baseline sample used by the Swiss Household Panel (called SHP_I) was interviewed for the first time in 1999. In 2004, we introduced a refreshment sample (called SHP_II) of addresses drawn up in accordance with the same sampling design as in the survey and started to interview a second panel. Surveys SHP_I and SHP_II are rigorously identical, the only difference being that in 2005, SHP_I was up to Wave 7 while SHP_II was on Wave 2. The SILC pilot survey in Switzerland was launched in 2004 for the years 2004-05, alongside SHP_II with the same sampling design. 60% of the questions asked in the surveys SHP and SILC pilot were identical or comparable.

The years 2004 and 2005 present very interesting possibilities for methodological studies:

1. Customary weights for SHP_I but also weights that enable SHP_I, SHP_II and SILC to be combined cross-sectionally were produced. The results from cross-sectional analyses could therefore be made more precise (larger sample).
2. Longitudinal weights combining two panels with different reference populations were produced.
3. In 2004, SHP_I was on Wave 6, whereas SHP_II and SILC pilot on Wave 1: the same questionnaires were administered both to the survival sample of 1999 and to the fresher sample of 2004. This fact was a unique occasion for us to conduct a very powerful attrition study and to identify the variables which are the more threatened. This provides on one side crucial information for validating and improving the non response models used in the weighting procedures and on the other side precious information for the data users.

The presentation will focus on the three quoted points.

7.7 Income Measurement in Surveys

Session coordinated by:

Vijay Verma; University of Siena, Italy

Martin Zelený; University of Economics, Prague, Czech Republic

Presentations

1. *From Survey to Social Indicators: Measurement and Editing of Income Data in the Austrian EU-SILC* – N. Lamei (Austria)
2. *Imputation in EU-SILC: an Application to Spain.* – M. Paniagua San Martín (Spain)
3. *From Sample Surveys to Totally Register-based Household Income Statistics: Experiences from Finland and Norway* – J. Epland (Norway), V. Törmälehto (Finland)
4. *The Income Question of the European Social Survey and a Proposal to Increase the Comparability of the Income Measurement in Social Surveys* – U. Warner (Luxembourg), J. Hoffmeyer-Zlotnik (Germany)
5. *Measuring Imputed Rent in EU-SILC Survey* – I. Masiulaityte (Lithuania), V. Balinskaitė (Lithuania)
6. *Fragility of Women's Wages in Luxembourg* – A. Haag (Luxembourg)
7. *Measuring Wages and Working Hours in a Worldwide Web-survey* – K. Tijdens (Netherlands)
8. *An Analysis of Monetary and Non-monetary Components of Households Economic Uneasiness* – M. Martini (Italy)
9. *Education Level, Unemployment, Gdp and Its Relationships: Experiences from the Czech Regions* – J. Fischer (Czech Republic), P. Mazouch (Czech Republic)

Presentations in the session "Income Measurement in Surveys"

From Survey to Social Indicators: Measurement and Editing of Income Data in the Austrian EU-SILC

Nadja Lamej; Statistics Austria, Austria

EU-SILC is a new European statistics in income and living conditions of people in private households. Its main focus is on the income situation of households, but other dimensions of living standard like material deprivation, health, education etc. are also thoroughly surveyed. Probably the most prominent indicator coming out of EU-SILC is the 'at risk-of-poverty rate', the percentage of people below a certain relative income threshold. Calculation of this highly political indicator depends on good quality data on household income, meaning that personal and household income components need to be at hand in full, reliably and accurately. Both the surveying as well as the post-survey treatment of those data therefore are critical for the final data quality and have to follow certain standards.

This presentation gives an insight into the processing of survey data in the Austrian EU-SILC survey: from the time a question is asked in a household and keyed into the CAPI-laptop to the final statistical analysis. EU-SILC has been conducted in Austria since 2003 and the requirement to each year produce consistent micro-datasets, fully checked, imputed and weighted has brought up numerous challenges so far. In describing the actual work process it is tried to convey how some of them were met. An overview over the complex system of data management for EU-SILC is provided and examples for good practices having so far been developed especially in the fields of checking the raw data and editing income-variables are given. Finally, in making this whole survey and data management process transparent it should become clearer for interested users and policy makers to see what lies behind the social indicators calculated from EU-SILC.

Imputation in EU-SILC: an Application to Spain.

Milagros Paniagua San Marfín; Instituto Nacional de Estadística (INE), Spain

The European Statistics on Income and Living Conditions (EU-SILC) survey's main target is to provide European countries with income data.

One of the well-known and inevitable phenomenon in surveys, in special when dealing with income, is the lack of response, what generates a bias and increases the variance.

There are two principal methods to correct the non-response: reweighting and imputation.

We distinguish three types of non-response: in a collaborating household we may find individuals giving partial information (partial non-response) and those who have some individual questionnaire lacking (individual non-response).

When the household does not give any answer to the questionnaires, we have a unit non-response.

In this paper, we will show how we treat the different non-response situations by imputing the missing values for partial and individual non-response. Specially, the imputation procedures, where IVE software is used.

Keywords: EU-SILC, income, non-response, imputation, IVEWare.

From Sample Surveys to Totally Register-based Household Income Statistics: Experiences from Finland and Norway

Jon Epland; Statistics Norway, Norway

Veli-Matti Törmälehto; Statistics Finland, Finland

With the launch of the European Union Statistics on Income and Living Conditions (EU-SILC) in 2003, all European countries now conduct an annual sample survey on income and living conditions in order

to deliver micro data with output-harmonised target variables on income, demography, labour, material deprivation, housing, and health for Eurostat and the European Commission.

The complete set of target variables required for EU-SILC cannot be collected without direct data collection from households. The flexible approach of EU-SILC permits, however, combined use of register and interview data. This combination method has been used in Norway and Finland to produce the national income distribution surveys (IDS). For the income data, EU-SILC is then simply a continuation and adaptation of national IDS practises.

With the available register information on income and demography, a totally register-based income statistics may also be produced in both countries. The advantages of compiling income distribution statistics from data which cover the whole population instead of a sample are obvious, especially for national users with greater demand for accuracy, small domain analyses, and geographical and longitudinal data than is required for EU-SILC.

The paper presents the experiences from Finland and Norway in producing income inequality and low income indicators from census-type sources. The main objective is to compare the key income inequality and low income indicators obtained from the total sources to those estimated from sample sources (IDS/EU-SILC), and to discuss the validity of the results.

The paper first discusses the extent such a totally register-based data would satisfy the internationally agreed concepts, e.g. those stated in the Canberra Group report and the EU-SILC regulation for household income statistics. The construction of appropriate income sharing unit is particularly relevant, the choices being the survey-defined economic households versus register-based dwelling units. The validity of income concepts needs discussion as well, especially in Finland where register income data still has some shortcomings. The common inequality and low income indicators from the Norwegian and the Finnish experiments are then presented and benchmarked with the sample-based sources.

The Income Question of the European Social Survey and a Proposal to Increase the Comparability of the Income Measurement in Social Surveys

Uwe Warner; Centre d'Etudes de Populations, de Pauvreté et de Politiques Socio-Economiques (CEPS / INSTEAD), Luxembourg

Jürgen H.P. Hoffmeyer-Zlotnik; Centre for Survey Research and Methodology (GESIS-ZUMA), Germany

During the last decade, the number of cross-national and cross-cultural empirical research has increased; at the same time the need for comparative survey data grew.

Four strategies of making data comparable are common:

1. the output harmonization - after the data collection by using different survey instruments in different nations, the comparative data are created;
2. the input harmonization - by using common fieldwork instruments and survey regulations comparative data are obtained;

Looking on the question the total net household income, we discuss the advantage and weakness of an input harmonized social survey. We demonstrate the impact of the national social, economic and legal particularities on the answering behavior of the surveyed respondent by comparing across countries the interview outcomes from the European Social Survey (ESS) and the European Community Household Panel (ECHP). ESS used a crude measurement of the total net household income interviewing only one randomly selected household member. ECHP surveyed all persons living in a sampled household and asked all income sources and components of the respondents and the household. In this paper we use ECHP as a reference and benchmark showing the most accurate method to measure income, and compare this with the interview results of ESS.

For comparative social surveys we propose a fieldwork instrument including a set of questions on income that takes into account the national circumstances. We get comparable survey data across countries reflecting the national specifics, the particular practices in the earning structures and the national habits in summing up the different income components. Such a new fieldwork instrument

integrated into the data production of cross-national surveys increases the analytical power of the comparative socio-demographic variable “total net household income”.

Measuring Imputed Rent in EU-SILC Survey

Inga Masiulaitė; Statistics Lithuania, Lithuania

Violeta Balinskaitė; Statistics Lithuania, Lithuania

The imputation of rental expenditures is an important step in the estimation of a household’s standard of living.

The imputed rent refers to the value that shall be imputed for all households which do not report paying full rent, either because they are owner-occupiers or because they live in accommodation rented at a lower price than the market price, or because the accommodation is provided them for free.

According studies that have been made with imputed rent in Household Budget survey (HBS) imputed rent in EU-SILC survey is analyzing. Some methods for imputation of the rent (like self-assessment, homogeneity groups, Heckman method and log-linear regression method) and additional information (data from Housing Rental Price Survey) are analyzed. Variables like location of the dwelling, number of rooms, type of dwelling and the amenities are taken into account in to the rent imputation. Most of the auxiliary variables correlate with the rent and make influence on the rent price.

Simulation is made with HBS data with different per cent of the tenants. Behaviour of the estimates of coefficient of variation of imputed rent estimated by different methods is analyzed. According the simulation results the best model for EU_SILC is selected.

Fragility of Women’s Wages in Luxembourg

Antoine Haag; Service Central de la Statistique et des Etudes Economiques (STATEC), Luxembourg

The multidecomposition of the Gini index is a powerful tool to compare wage disparities. This index obtained is more complete than its one-dimensional configuration. This article shows a decomposition of the Gini index into sub-groups in order to point out disparities by gender,

occupations, activity sectors, ages and nationalities. For example, the exploitation of the Labour Force Survey reveals that in 2001 (resp. 2005) the monthly net gap in net wages is of 714 euros (resp. 808 EUR) for two men randomly drawn out of the population (individuals working full time) and of 627 euros (resp. 702 EUR) for two women. Opposite to some other indicators’ families (particularly those issued from entropy). the proposed index decomposition can explain the overlapping degree of the two distribution curves. Variance and asymmetry phenomena are being taken into account with this approach. Female wage profiles tend to be more male driven while still showing disparities in ages, nationalities or occupations.

Keywords : Decomposition of the Gini index, Wage inequalities, Entropy, Multi-decomposition.

Measuring Wages and Working Hours in a Worldwide Web-survey

Kea Tijdens; University of Amsterdam, Netherlands

The WageIndicator is an international, continuous, volunteer web-survey on work and wages in 19 countries. The web-survey is available at frequently visited WageIndicator websites, which jointly received almost 8 million unique visitors in 2006. The Salary Checker on the web-sites is the major attraction. Worldwide, people show a great desire for information about wages, as the numbers of visitors as well as their emails show. The websites are deeply rooted in the Internet. Many other websites link to WageIndicator websites. They are easily found in search engines. They have a well developed marketing strategy including cooperation with major players on the Internet. See www.wageindicator.org/.

All websites in all countries use the same web-survey in their native language(s). The target population of the web-survey is the labour force, including job seekers and the informal economy. Distinct

groups within the target population route differently through the questionnaire. Parallel questions are used to address specific groups. The web-survey is divided into six sections: (1) occupation and education; (2) workplace characteristics; (3) employment history; (4) working hours; (5) employment contract and salary; (6) personal questions. Each section ends with attitudes and opinions regarding the topics addressed. The web-survey employs search trees for the questions 'What is your occupation' or 'In what industry you do work', allowing for detailed coding according to the international classifications ISCO and NACE respectively. It takes on average 20 minutes to complete the questionnaire.

This paper first reviews the income questions in four large-scale European surveys. Next, it details the income survey questions and their routing in the WageIndicator web-survey. It explains what considerations have been taken into account as for measuring gross or net wages, payment period, income earned by self-employed, issues related to payment in kind, payment modes (cash/bank account), bonuses, etceteras. The data, including an open-ended question to comment upon the reported wages, shows particularly that in all countries under study the payment period reveals much more variation than initially expected, and than is shown in the survey questions of the four large-scale European surveys, who force the respondent into a payment period category, e.g. monthly or weekly earnings. Third, the paper details how working hours are measured. Fourth, the paper details how hourly wages are calculated. Finally, the paper details the plans for instant calculation of hourly wages in the web-survey, testing for reliable minimum and maximum values. If the respondent does not pass the test, the subsequent pop-ups to be used for controlling the reported working hours, wages and pay periods are discussed.

An Analysis of Monetary and Non-monetary Components of Households Economic Uneasiness

Maria Cristiana Martini; University of Padua, Italy

Poverty is not merely a monetary matter, but a multidimensional concept. Income is only one of the several components which determine the economic problems but, apart from extreme situations, in general it is not a sufficient cause of uneasiness. Households with equal income levels can express different degrees of economic sufferance depending on their expectations and living standards, as well as on the familiar structure, the household characteristics or other specific problems which may affect the individual members.

This contribution aims at analysing the causes of familiar frailty which, together with a low income, contribute to trigger the economic uneasiness. The data come from a survey on households living in the Veneto Region, which has been organised by the Local Authority for Social Policies, Charity and Non Profit, the Centre for Households Documentation and Analysis and the Regional Health Service, together with the University of Padua; the survey was conducted on a sample of 2459 households (in total 6920 individual member) by means of a CATI (Computer-Assisted Telephone Interview) methodology.

The analyses outline the existence of familiar structures which are particularly liable for economic uneasiness, as single parent families, newly formed young families with children or elderly households. Moreover, the contemporaneous presence of health or relational problems, or some contingent events like unexpected expenses, job troubles or housing difficulties can amplify the importance of the economic uneasiness. Furthermore, we identify different typologies of households with economic problems, in order to describe different profiles of households at risk, that is, different 'poverties' which should be coped with by means of different strategies of economic support and customised social policies.

Education Level, Unemployment, Gdp and Its Relationships: Experiences from the Czech Regions

Jakub Fischer; University of Economics, Prague, Czech Republic

Peter Mazouch; University of Economics, Prague, Czech Republic

Detecting the relation between unemployment and business cycle (or economic level) is a classical issue for economists and is a subject of many papers. In our opinion it is necessary to take into account

the structure of the unemployment, when the unemployment rate is broken down by the education level. Our paper is focused not only on the analysis of trends of total unemployment rate, but also on trends of specific rates and their relation with GDP increase/decrease. Education-specific rates of unemployment are very different in the Czech Republic (the highest rate is for lower-educated people), but the sensitivity on the business cycle is also significant. The other part of this paper is focused on the relationship of the education-specific rate of unemployment with the regionalized gross domestic product (RGDP) per capita in given regions. Relationship between unemployment rates and RGDP signalizes, that in regions with the higher level of value added the rate of unemployment is smaller and this relation is true for all the education-specific groups and the sensitivity of unemployment rate in individual groups decreases with a rising education level.

7.8 Survey Measurement of Service Quality

Session coordinated by:

Irena Ograjenšek; University of Ljubljana, Slovenia

Presentations

1. *Measurement of Service Quality: Benchmarking Surveys to Other Applicable Measurement Tools* – I. Ograjenšek (Slovenia)
2. *Delineating Customer Satisfaction from Perceived Service Quality at the Level of a Tourist Destination* – T. Dmitrović (Slovenia), L. Knežević Cvelbar (Slovenia), M. Makovec Brenčič (Slovenia), I. Ograjenšek (Slovenia), V. Zabkar (Slovenia)
3. *Validation Procedure and the Validity of Service Quality Measurement Tool: the Case of Tourism Services* – B. Milfelner (Slovenia), D. Mumel (Slovenia), B. Snoj (Slovenia), A. Pisnik Korda (Slovenia)
4. *A Tale of Two (or More) Surveys: the Application of a Latent Variable Modeling for Linking Attributes on Separate Conjoint Surveys* – J. Dang (United States), J. Cole (United States), J. Magidson (United States)
5. *An Applet to Measure and to Assess the Customer Satisfaction: How Neo Graduate People in the Working World Evaluate Their University Syllabus* – E. Zavarrone (Italy), M. Civardi (Italy)

Presentations in the session "Survey Measurement of Service Quality"

Measurement of Service Quality: Benchmarking Surveys to Other Applicable Measurement Tools

Irena Ograjenšek; University of Ljubljana, Slovenia

There are two distinct approaches to the use of statistical methods in continuous quality improvement of services. These are the production approach, which uses the traditional statistical quality control tools, and the marketing approach, which builds on the social sciences toolbox (especially surveys).

Production approach focuses on optimisation of service processes from the management viewpoint. Quality standards used in this framework can be expressed in non-monetary (e.g. the tolerance level for the monthly number of customer complaints, the health and hygiene rules for workers, the number of customers served in a given time frame) or monetary terms (expressed as upper tolerance levels for different categories of quality costs), and can be applied in observational as well as inferential studies using tools such as control charts or methods such as design of experiments.

Marketing approach focuses on the customer assessment of service experience, using surveys as a predominant tool. Emphasis is on analysis of separate quality attributes as well as overall quality. Like in the production approach, quality standards in the framework of the marketing approach can be expressed in non-monetary or monetary terms. When defining quality standards in non-monetary terms, there is little room to avoid the highly subjective customer expectations formed through search, experience and belief (or credence) attributes. Customer intentions can also be used as subjective non-monetary quality standards. This is problematic when they reflect the ideal rather than actual buying behaviour of customers who act as quality evaluators.

In this paper we compare and contrast the production and marketing approach in detail, summarizing the most important advantages and disadvantages of surveys in comparison with measurement tools applicable in the framework of the production approach.

Delineating Customer Satisfaction from Perceived Service Quality at the Level of a Tourist Destination

Tanja Dmitrović; University of Ljubljana, Slovenia

Ljubica Knežević Cvelbar; University of Ljubljana, Slovenia

Maja Makovec Brenčič; University of Ljubljana, Slovenia

Irena Ograjenšek; University of Ljubljana, Slovenia

Vesna Zabkar; University of Ljubljana, Slovenia

Tourist destination can be defined as a bundle of tourist products and services consumed under the same brand name, thus offering the consumer an integrated experience. Traditionally, destinations are described as well-defined geographical areas such as countries, islands, or towns, which their visitors perceive as a unique entity (Buhalis, 2000).

Understanding customer satisfaction is without doubt one of the most important competitive advantages for companies operating in the tourism sector. Increasing tourist satisfaction at the level of destination could improve destination's competitive positioning and facilitate understanding of which factors are creating higher attractiveness of destinations. It is therefore important for successful destination marketing because it influences the choice of destination and the decision to return. Different supply-side stakeholders should therefore strive towards one common goal: maximization of customer satisfaction.

Literature research shows that two of the main problems accompanying the modelling of customer satisfaction are defining and operationalising customer satisfaction as opposed to perceived service quality. Although research has suggested both that perceived service quality and customer satisfaction are

distinct constructs (Oliver, 1997; Taylor and Baker, 1994), and that there is a causal relationship between the two (Cronin and Taylor, 1992; Gotlieb et al., 1994; Spreng and Mackoy, 1996), the constructs have in some cases been used interchangeably (Iacobucci et al., 1994; Parasuraman et al., 1994; Taylor and Baker, 1994; Oliver, 1997; Mittal et al., 1998).

In this paper we strive to delineate the two constructs from one another and determine the direction of the causal relationship between them at the level of a tourist destination.

Validation Procedure and the Validity of Service Quality Measurement Tool: the Case of Tourism Services

Borut Milfelner; University of Maribor, Slovenia

Damijan Mumel; University of Maribor, Slovenia

Boris Snoj; University of Maribor, Slovenia

Aleksandra Pisnik Korda; University of Maribor, Slovenia

In the last decade, a remarkable development in the study of quality perceptions has been made possible due to the emergence of diverse measurement instruments (e.g. SERVQUAL, SERVPERF). The existence of these instruments has served as the basis of development of an important research agenda oriented towards the application of measurement of service quality and its antecedents. However, without taking into consideration the specifics of the industry, the measurement methodology of service quality can not be applied in the marketing practice. This also holds true for tourism which is known for its complex and diverse range of products and services.

The utility of service quality measurement methodology depends on the accurate functioning of the measurement instruments on which the investigations are based and these instruments' validity and reliability. Inadequate instruments and misidentified variables are too often the consequence of incorrect application of 'universal' scales to specific research areas.

According to this, the authors present the development and justification process of the instrument for measuring tourism service quality in the Slovenian context. Proposed measurement instrument for tourism service quality is similar to the SERFPERF model (Cronin & Taylor, 1994), since such method can account for most of the variance in the whole quality construct. Content validity of the instrument is assessed by conducting several interviews with experts in the field, and face validity with the pre-tests by the members of the target population (focus groups). Additionally, different viewpoints of construct validity are assessed in order to test the extent to which a measure behaves in the theoretically sound manner. Convergent validity is assessed with exploratory factor analysis (EFA), discriminant validity with confirmatory factor analysis (CFA), and nomological validity with regression analysis. Results of tests for reliability of the construct such as Cronbach's alpha, composite reliability, and average variance extracted, are also provided.

A Tale of Two (or More) Surveys: the Application of a Latent Variable Modeling for Linking Attributes on Separate Conjoint Surveys

Jeff Dang; Consulting Measurement Group, Inc., United States

Jason Cole; QualityMetric, United States

Jay Magidson; Consulting Measurement Group, Inc., United States

Conjoint analysis is a rigorous survey technique that has been increasingly used in academic and commercial settings to investigate people's preferences. In a traditional conjoint study, respondents are presented with a full profile of all of the attributes (or domains) and features (or levels) for a particular product or service. However, research involving a large number of attributes can be too complex for some respondents, too burdensome for many respondents, and has been shown to elicit inaccurate responses. The current study describes a procedure used to link attributes from two or more different conjoint surveys that share at least one attribute. This novel method can help prevent information

overload and lessen the burden on respondents while allowing utility parameters to be estimated on a large number of attributes.

Choice based data were linked using a partial profile design and parameters were calculated using maximum likelihood estimation for finite mixture modeling. Several examples demonstrating the procedures used to link choice based survey data are provided. In addition, results from a latent variable modeling of the linked survey data will be reviewed. Finally, to illustrate the flexibility of latent conjoint analysis, continuous and categorical covariates were simultaneously estimated to demonstrate the usefulness of covariate based profiling.

An Applet to Measure and to Assess the Customer Satisfaction: How Neo Graduate People in the Working World Evaluate Their University Syllabus

Emma Zavarrone; Università di Milano-Bicocca, Italy

Marisa Civardi; Università di Milano-Bicocca, Italy

In this paper we study the properties of a family of index, called CI. These indices have been proposed by Civardi, Zavarrone (2003). In order to evaluate quality service in consumer field, the most frequently used scales offer four or five points and the first two (or the last two) points on both scales are associated with negative evaluations and the last two (or the first two) are associated with symmetric positive evaluations. To have an immediate and computationally not expensive application of the CI the authors present a software for the evaluation of the satisfaction, given a number of answers (questionnaires).

The final part of the paper is devoted to a practical application of the CI on data collected on the labour survey of July 2005. This survey STELLA1 was made up by eight Italian Universities on its former students with a bachelor taken at the third quarter 2003 and the first quarter 2004.

8 Specific Topics II

Sessions in Specific Topics II

Ethical Considerations in Survey Research	226
Secularisation Under Different Regimes	230
Health Surveys in Europe	233
The Stability of Political Opinions & Attitudes	236
Time Use Surveys	240
Measuring and Explaining the Fear of Crime	247
Survey Data and Public Dissemination	250
Survey Evaluation Methods Using Behaviour Coding and Sequence Analysis	254
Difficult Groups in Survey Research	257

8.1 Ethical Considerations in Survey Research

Rebecca Taylor; National Centre for Social Research - NatCen, United Kingdom

Gerry Nicolaas; National Centre for Social Research - NatCen, United Kingdom

Presentations

1. *Doing the Right Thing: Privacy and the Uk Census* – C. Heeney (United Kingdom)
2. *Ethics Around the Participation of Children and Young People in Survey Research: Listening to the Perspectives of Children and Young People* – A. Bell (United Kingdom), C. Bryson (United Kingdom), J. Graham (United Kingdom), R. Ormston (United Kingdom), C. White (United Kingdom)
3. *Ethical Review of Public Health Research in England* – B. Erens (United Kingdom)

Presentations in the session “Ethical Considerations in Survey Research”

Doing the Right Thing: Privacy and the Uk Census

Catherine Heeney; University of Oxford, United Kingdom

National Statistical Institutes (NSIs) are the organisations with responsibility for compiling official statistics at a national level. This paper discusses both the ethical and legal context for the work of NSIs and reflects upon data gathered at the time of the 2001 UK Census. Using the results of 30 semi-structured interviews with a cross section of the public the paper will consider the relationship between values, experiences and beliefs, the official governance framework and the actual uses of statistical data. Public support for the work of NSIs, which includes carrying out a Census, is often sought on the basis that society as a whole will benefit from individual participation, while the individual themselves will not suffer direct negative consequences. This situation resonates with what liberal theorist John Rawls refers to as the ‘contract tradition’. Similar justifications can be found in the official governance framework for the production of official statistics and supporting documents. The paper will discuss the various pieces of legislation and guidelines that constitute the official governance framework for NSIs, including the UK Data Protection (1998) and the European Statistics Code of Practice (2005). It will consider how ethical values are employed in laws and guidance that govern the work of NSIs and whether this produces an appropriate understanding of the role of official statistics. The paper argues that an appropriate understanding must also take into account the particular character of today’s information society, where there is a demand for statistical data for a wide variety of purposes and where tools for data analysis can pick out subpopulations with increasing sophistication.

One of the benefits often-cited in ‘grey’ literature describing the work of NSIs is that statistical information allows for an enlightened planning of services and resource allocation based on the needs of the populace. NSIs aim to strike a balance between ensuring that the data sets they assemble are available as a resource for society (where ‘society’ can include private sector agencies) and the privacy of individual data subjects (understood as closely related to the protection of their identities). The paper will consider the implications of the current situation with regard to values, beliefs and the realities of data processing. This will include reflection on the secondary uses of data from the Census and the issue of maintaining trust in the exercise of Official Statistics. It will be argued that there is a disjuncture between the aims of the official governance framework and the outcomes of uses of statistical data. As a result participants separate an ‘ideal statistical situation’ which reflects the normative stance taken by law and guidance, and a murkier though not fully understood reality. Arguably if a combination of data can be used in profiling groups and areas this has the potential to affect individual interests in a direct way. Such practices as the targeting or exclusion of particular groups and their members are enabled by innovations in information technology. This challenges the standpoint taken in the official governance framework for the use of statistics. This is a grey area for governance and potentially damaging for public support of activities such as the Census.

Ethics Around the Participation of Children and Young People in Survey Research: Listening to the Perspectives of Children and Young People

Alice Bell; National Centre for Social Research - NatCen, United Kingdom

Caroline Bryson; National Centre for Social Research - NatCen, United Kingdom

Jenny Graham; National Centre for Social Research - NatCen, United Kingdom

Rachel Ormston; National Centre for Social Research - NatCen, United Kingdom

Clarissa White; National Centre for Social Research - NatCen, United Kingdom

We propose a paper that presents recent findings from a small-scale qualitative study among children and young people aged between 7 and 15, which explored their views about taking part in surveys.

The study adds to a currently small evidence base about the perspectives of children and young people about the ethics of taking part in survey research.

Increasing emphasis on ‘listening to children’s voices’ about matters that affect them together with burgeoning debates around issues of ethics in social research have resulted in a number of texts explicitly concerned with issues of ethics as they relate to research with children and young people. However, there is very little evidence about the perspectives of children themselves on questions of research ethics. Moreover, what evidence there is tends to focus on qualitative research, which can deploy a greater degree of flexibility or adaptability in its approach than quantitative research, which requires the need for a representative sample, a relatively inflexible data collection instrument and the employment and training a much larger number of interviewers.

Our study involved a series of eight projective discussion groups with children in primary and secondary schools. We used a video depicting hypothetical scenarios as our main stimulus tool. This innovative approach (rather than following up on respondents from a particular study) enabled us to address issues in the absence of any influences stemming from the particular characteristics of a single research experience.

Our paper would present evidence from our study about the perspectives of children and young people on –

- Methods and approaches for obtaining informed consent from children and young people;
- The role of parents and other third parties in providing informed consent;
- How a child might come to provide or withhold informed consent and factors influencing the decision-making process;
- Attitudes towards privacy during the interview and factors influencing the views of children and young people (eg topic, location, interviewer characteristics);
- Coping with discomfort and withdrawal during the interview (strategies, feasibility, role of the interviewer, reasons);
- Attitudes to confidentiality, anonymity and disclosure;
- Factors that might lead a child or young person to withdraw from research or withdraw data after the interview – and ways of facilitating a child’s right to exercise choice in this area.

Ethical Review of Public Health Research in England

Bob Erens; National Centre for Social Research - NatCen, United Kingdom

Despite having more in common with social surveys than traditional medical research (e.g. RCTs), public health surveys and patient surveys in England have traditionally been subjected to ethical review by medical research ethics committees (RECs). Ethical review is now being routinely extended to many other areas of social research, and there are important lessons to be learned from the experience of health and patient surveys being reviewed by medical RECs.

According to official guidance, it is explicitly not the task of RECs to consider the ‘science’ of the project, as the study should have been peer reviewed before submission to the REC. What RECs should consider are: risks and benefits to participants; recruitment procedures; obtaining informed consent; maintaining confidentiality; the study design being appropriate to answer the research questions; and the relevance of the research to the community.

However, on a number of recent health/patient surveys in England, medical RECs have attempted to impose changes on the research design that could have seriously compromised the integrity of the studies. Some of the issues of concern addressed in the paper include:

- Consistency, both between different RECs, and within the same REC over time. Submitting a REC application can be a lottery, as some RECs will allow practices that other RECs will not.

- Confidentiality, with some RECs insisting that survey researchers transgress their pledges of confidentiality by informing on respondents (e.g. in possible cases of child abuse) or by releasing data to third parties (e.g. hospital managers if possible 'malpractice' by medical staff has been mentioned).
- Obtaining informed consent, which is often the most contentious issue with medical RECs. This can vary from the seemingly trivial (e.g. refusing the use of the word 'important' to refer to a survey in an advance letter because its considered 'coercive'), to quite fundamental design issues likely to seriously affect response rates (e.g. insisting on an opt-in rather than the more commonly used opt-out procedure).

The problems of medical RECs considering public health/patient surveys stem from: a) REC members having little awareness/knowledge of standard social survey procedures and ethical guidelines; and b) the fact that RECs were established mainly to consider clinical trials, which have a very different approach to obtaining informed consent than do social surveys. While its undoubtedly proper that social surveys are increasingly subjected to independent ethical scrutiny, its important that the principles of ethical research are interpreted in such a way that reflects accepted survey procedures rather than those devised largely for traditional clinical/pharmaceutical research.

8.2 Secularisation Under Different Regimes

Session coordinated by:

Heiner Meulemann; University of Köln, Germany

Presentations

1. *European Religiosity at the Turn of the Millennium: the Impact of Regional Characteristics* – L. Halman (Netherlands), J. Gelissen (Netherlands)
2. *Religious Change 1981- 2006: Decline, Stability or Increase? a Study of the Differential Effects of Human Security and Cultural Diversity* – T. Pettersson (Sweden)
3. *Religious Attendance in Cross-national and Longitudinal Perspective: a Test of Rival Theories* – S. Ruiter (Netherlands)
4. *Religiosity East and West* – A. Kawabata (Japan)
5. *Secularization in Japan from a Comparative Perspective* – K. Manabe (Japan)
6. *Changing Godly Images in Europe* – N. Moor (Netherlands), A. Need (Netherlands), W. Ultee (Netherlands)
7. *Sociological Determinants of Atheism in Secularized Europe* – O. Riis (Denmark)
8. *Religion, Secularization and Well-being* – H. Meulemann (Germany)

Presentations in the session "Secularisation Under Different Regimes"

European Religiosity at the Turn of the Millennium: the Impact of Regional Characteristics

Loek Halman; University of Tilburg, Netherlands

John Gelissen; University of Tilburg, Netherlands

In a recent article (Halman & Draulans, 2006), the degree to which European people are secular was investigated, focusing not only on religious practices, but also on beliefs. Data from the recent European Values Study surveys revealed that religious denomination as well as cultural and socio-economic heritages explain part of the patchwork pattern in levels of religiosity and religious participation in contemporary Europe. It was also concluded that the individual level hypotheses were more strongly corroborated than the country level hypotheses, and that a substantial amount of individual level variance in religious practices and beliefs remained to be explained.

Most studies on religious beliefs and behaviour thus far have focused on the extent to which individual and country characteristics contribute to the explanation of differences in individual religious beliefs and behaviour. However, research, which additionally includes regional differences within nation-states as potential explanations of individual beliefs and behavior, is lacking thus far. The inclusion of such regional characteristics may be important for several reasons. For example, important regional structural disparities exist with respect to wealth, education and age-composition and such differences may correlate with religiosity. Moreover, Hearl, Budge, and Pearson (1996) point to important cultural differences among regions and these differences may also systematically relate to people's religious beliefs and behavior. In a multi-level design we explore the impact of individual characteristics and regional features on three religious dimensions: religious beliefs and religious values and religious practices.

Religious Change 1981- 2006: Decline, Stability or Increase? a Study of the Differential Effects of Human Security and Cultural Diversity

Thorleif Pettersson; University of Uppsala, Sweden

Religious Attendance in Cross-national and Longitudinal Perspective: a Test of Rival Theories

Stijn Ruiter; Radboud University Nijmegen, Netherlands

This article uses the secularization paradigm and the religious market paradigm to explain cross-national and longitudinal variations in religious participation. From the religious market paradigm, it derives the religious regulation theory. From the secularization paradigm, it derives theories on the modernization of ideologies, social ties, and economies. The theories are tested simultaneously with pooled data for 60 nations, obtained from the World Values Surveys (1990-2001) and the International Social Survey Programme (1991-1998). The results from the multilevel methods do not support the religious regulation theory and the theory on the modernization of ideologies. There is strong evidence, however, for the theories on the modernization of social ties and economies. It is discussed how these theories explain that religious attendance remains stable in some nations, whereas it increases or decreases in other countries.

Religiosity East and West

Akira Kawabata; University of Osaka, Japan

Secularization in Japan from a Comparative Perspective

Kazufumi Manabe; Kwansai Gakuin University, Japan

Changing Godly Images in Europe

Nienke Moor; Radboud University Nijmegen, Netherlands

Ariana Need; Radboud University Nijmegen, Netherlands

Wout Ultee; Radboud University Nijmegen, Netherlands

In the sociology of religion secularization is often considered identical to a drop in church attendance. Adherents of the modernization theory show that over time people less often visit church and less often consider themselves church members. These research findings relate to religious rites. Like Durkheim already stated in 1912, religion consists of two fundamental categories: beliefs and rites. In this article we contribute to the sociology of religion by studying and explaining the trend in godly images.

According to Lenski's ecological evolutionism (1970) godly images are a byproduct of technological advancement. Lenski hereby states that some religious ideas, which were perfectly understandable for people in the agrarian societies, appear unfamiliar to people in industrial societies. He claims that in modern societies the image of a supreme powerful god no longer corresponds with the idea that people take control of their own life. Because the political power in agrarian societies is more personal than the bureaucratic power in industrial societies, man's idea of the ultimate power is expected to have changed. We try to explain the trend in godly images with both individual characteristics and societal characteristics in the period in which people were socialized. We pay extra attention to the industrial and political history of societies.

We study changes in godly images by comparing the beliefs from people in different birth cohorts. In our research we focus on people from European countries where Christianity is the largest and most important religion. We make use of the European Values Studies and study religious change by centering our attention to the image people from different birth cohorts have of their god. We differentiate between a personal god and a more abstract god (spirit or life force). We expect that religious beliefs in industrializing societies are becoming less personal.

Sociological Determinants of Atheism in Secularized Europe

Ole Riis; University of Copenhagen, Denmark

Religion, Secularization and Well-being

Heiner Meulemann; University of Köln, Germany

Well-being depends on self-enhancement and self-transcendence on the level of persons and on the development of countries. Self-enhancement is measured by success in the life careers, self-transcendence by religiosity and altruistic attitudes. In particular, it is examined whether religiosity increases well-being once the other person level variables have been controlled for. The development of countries is measured economically, politically, and societally. In particular, it is examined whether the development of the civic society increases, and the degree of secularization decreases mean well-being of countries, once person level variables have been controlled for. Dependent variables are the self-assessed life-satisfaction and happiness, data the European Social Survey 2002 and 2004 with altogether 48 country*time samples of 88 040 respondents. To examine person level and country level effects, intercept models of multi-level regressions are applied.

8.3 Health Surveys in Europe

Session coordinated by:

Cornelia Bormann; University of Applied Sciences, Bielefeld, Germany

Presentations

1. *The German Health Monitoring System: Options for Longitudinal Modules* – T. Ziese (Germany), C. Lange (Germany), B. Kurth (Germany), . Scheidt-Nave (Germany)
2. *Extending Survey Based Research with Routine Administrative Data in Healthcare Research* – G. Heller (Germany)
3. *The Relationship Between Housing Conditions and Health - Some Findings from the WHO LARES Study of 8 European Cities* – S. Nicols (United Kingdom)

Presentations in the session "Health Surveys in Europe"

The German Health Monitoring System: Options for Longitudinal Modules

Thomas Ziese; Robert Koch Institute, Germany

C. Lange; Robert Koch Institute, Germany

BM Kurth; Robert Koch Institute, Germany

Scheidt-Nave; Robert Koch Institute, Germany

The Robert Koch Institute (RKI) has conducted several health surveys on irregular intervals since 1984 on national level. Up to the reunification of the country, two national health interview and examination surveys (HIS/HES) were conducted in former West Germany in 1984/86 and 1987/89. In 1990/91, 25-69-year old citizens of former West and East Germany were surveyed in order to compare health status and health-related behaviour between the two parts of Germany.

Two HIS/HES have been conducted since then: a survey of adults 18-79 years in 1998 (n>7000), and a survey of children and adolescents 0-17 years in 2003-2006 (n>17 000). Additionally national health interview surveys of the adult population (n>6000) have been carried out annually since 2003.

From 2008 on the Robert Koch-Institute plans to implement a continuous health monitoring system based primarily on two types of regular cross-sectional health surveys: Annual health interviews surveys and perennial health examination surveys. It is intended to integrate additional longitudinal elements into this monitoring system:

1. The participants of the HIS/HES from 1998 will be asked to participate in follow-up surveys in order to transform the cross sectional approach into a panel for adults.
2. The survey on children and adolescents from 2003-2006 shall be the basis for a cohort study to monitor changes in behaviour, conditions and outcomes from childhood on.

By combining cross-sectional surveys with cohort studies the German health monitoring system will provide valuable data for the German health reporting system, for health policy advising, the information of the public, and health related research. Examples of particularly relevant data include: trends in the prevalence of specific health conditions and health-related behaviour; health care utilization and identification of specific health risks groups (e. g. migrants or socially disadvantaged persons).

Extending Survey Based Research with Routine Administrative Data in Healthcare Research

Günther Heller; Research Institute of the Local Healthcare Insurance (WIdO), Germany

From a general perspective survey based research has several advantages, but also some notable drawbacks:

- One important limitation is that usually only samples instead of populations are included in a survey. Although sampling techniques may solve this problems in some instances, it is common sense that numerous research questions remain unanswered, especially when addressing research questions focussing using past data (classical retrospective designs) or longitudinal data.
- Another limitation is that validation of data quality is implemented only in a few surveys and usually restricted to limited samples.

Both problems may be addressed using routine administrative data as these data usually covers complete populations and is validated on a routine basis for other reasons: E. g. Hospital Episode Data for reimbursement of hospitals.

As a consequence recent research as well as legislative regulations have emphasized the combination of survey methodology with data from routine administrative sources. The Social Insurance Code § 303 ff. e. g. describes future usage of health insurance data for in scientific settings in Germany.

In this presentation we describe the contents of health insurance data according to the Social Insurance Code § 303 ff and provide examples how to extend survey based research with routine administrative data in healthcare research.

The Relationship Between Housing Conditions and Health - Some Findings from the WHO LARES Study of 8 European Cities

Simon Nicols; Housing Centre, Building Research Establishment, United Kingdom

The WHO LARES survey collected data on the relationship between housing and health in 8 European cities in 2003/04. This paper uses data from the survey to examine the relationship between housing (in particular damp housing) and the health of the occupants.

The participating cities were: Angers (France), Bonn (Germany), Bratislava (Slovakia), Budapest (Hungary), Ferreira (Portugal), Forli (Italy), Geneva (Switzerland) and Vilnius (Lithuania). Some 8,500 people were interviewed regarding their health and circumstances, and a physical survey of some 3,300 homes was undertaken by a trained inspector. The cities were self selecting and were not to be regarded as representative of their countries. But the data set could be used to draw comparisons between the cities, as well as being used as a whole to look at general relationships between housing and health.

The levels of dampness and mould growth varied between the cities, depending on the state of repair of the homes, the heating systems present, and internal and external temperatures.

The study identified a definite relationship between damp/mouldy homes and: anxiety/depression and migraine/frequent headaches from the group of chronic illnesses; diarrhoea and cold/throat illnesses from the group of acute illnesses; and asthma, wheezing, eczema, watery eyes/eye inflammation, headaches from the list of symptoms. This relationship does not imply cause and effect. Many illnesses appear to be mental conditions, and even the physical symptoms are of the sort which could be regarded as being the emotional response to circumstances – such as feeling trapped in poor housing.

8.4 The Stability of Political Opinions & Attitudes

Session coordinated by:

William van der Veld; University of Amsterdam, Netherlands

Presentations

1. *Attitude Stability and the Recollection of Past Attitudes* – E. Jaspers (Netherlands), M. Lubbers (Netherlands)
2. *Stability and Change in Party Preference in Switzerland* – U. Kuhn (Switzerland), B. Wernli (Switzerland)
3. *Alleviation of Context Effects in Attitude Questions Through Computer Assisted Interviewing and a Planned Missing Data Design* – L. Littvay (Hungary), C. Dawes (United States)
4. *Effect of Personal Experience on Attitudes: a Path Analysis with Panel Data* – A. Tutkun (Switzerland)
5. *Issue Ownership and Stability and Change in Party Choice* – S. Nicolet (Switzerland), P. Sciarini (Switzerland)

Presentations in the session "The Stability of Political Opinions & Attitudes"

Attitude Stability and the Recollection of Past Attitudes

Eva Jaspers; Radboud University Nijmegen, Netherlands

Marcel Lubbers; Radboud University Nijmegen, Netherlands

Many Western societies have experienced changes in attitudes on various moral issues, such as homosexuality and euthanasia. The Netherlands has seen a dramatic decline in opposition towards these issues. The attitude toward ethnic minority members has become much more negative over the same time period. The changes in the aggregate attitudes are often interpreted in terms of the changing composition of society. However, periodical circumstances might induce attitude change within individuals. We choose issues that have been debated intensely over the last half of the twentieth century, in order to achieve maximum saliency of the attitudes for all respondents. We explore the changes in attitudes people experience as they age, and whether and how they perceive this change. We use measurements on three different attitudes, to perform a more extensive analysis on attitude stability.

In the initial survey, over 2,000 individuals participated. We were able to find the current addresses of roughly 1,300 living respondents, aged thirty to eighty years. The successfulness of the second wave of the survey, was partly dependent on cooperation of the Dutch municipalities to help trace the addresses of respondents that had moved. Of those who received the second questionnaire, over 55 percent participated, resulting in a dataset of 854 respondents.

We investigate who is better in recalling past attitudes, but will also use both the recalled and the factual past attitudes for an analysis on attitude change. The first analysis allows us to test how reliable these recalled attitude measurements are. The second analysis will show how useful (or not) the retrospective attitudinal data are for determining predictors of attitude change. Preliminary results show that recalled attitudes are – strongly- biased toward the present attitude, but that some valid inferences on attitude change can be made when using the retrospective attitude measurements instead of the original 1995 data.

We find very different stability rates, depending on both the attitude we study and the way it was originally measured. Those who changed have a harder time recollecting their previous attitude toward euthanasia, but not toward for instance homosexuals.

Stability and Change in Party Preference in Switzerland

Ursina Kuhn; University of Neuchâtel, Switzerland

Boris Wernli; University of Neuchâtel, Switzerland

We look at changes in individual party preference between 1999 and 2005 in Switzerland using data from the Swiss Household Panel. Party preference is measured annually by the party one would vote for if there was a parliamentary election tomorrow. Considering only respondents with the right to vote, there are 2949 individuals who participated in all waves and between 4446 and 6740 that participated in the single waves.

First we discuss the quality of the data, theoretical concepts and past research.

Second we describe patterns of change in party preference in Switzerland. Among the respondents that participated in 7 waves, only 21 % had a stable party preference. 32 % changed between one particular party and having no party preference. 14 % never indicated a party preference in seven years and 32 % changed at least once from one party to another. The stability is overestimated due to truncated data and attrition. Even though there is little comparison with other data, the stability of party preference in Switzerland seems to be rather low. This can be explained by the large number of parties and the lower salience of national elections in Switzerland because of direct democracy and federalism. Looking at transitions between two consecutive years, we observe that respondents are more likely to indicate a party preference in election years than in other years. For the election in 2003,

there are considerable campaign-effects: compared to its support before the election, the Christian Democratic party succeeded in diminishing its losses while the losses of the Liberal party increased in the year before the elections.

Third we analyze how respondents with stable and changing party preference differ both generally and respective to the main parties. For this, we analyze transitions between elections in contrasting the answers of two waves in the Panel and by pooling data. First, we compare respondents with stable party preference for any party with those who change between having and not having a party preference (mobilized and demobilized voters). Then we look at the main parties separately, in order to find out how stable voters for a party (1) differ from new voters (2) and lost voters (3) of those parties, by applying multinomial logistic regression. We test demographic, socio-economic, and different political variables. Generally, we see that stable parties have more extreme or typical characteristics than occasional voters for that party. We can also see who is mobilized or demobilized by which party, which issues are salient for which party to attract new voters and how political satisfaction influences the changing behavior in party preference.

Alleviation of Context Effects in Attitude Questions Through Computer Assisted Interviewing and a Planned Missing Data Design

Levente Littvay; Central European University, Hungary

Christopher T. Dawes; University of California, San Diego, United States

This study proposes a data collection design and analytical method that could alleviate some context effects in attitude questions. Treating context effects as measurement error data can be analyzed using confirmatory factor analysis or item response theory where multiple questions measuring the same attitude are used as indicators/items of a latent construct. This latent construct is a continuous measurement error (i.e. context) free realization of the attitude measured and can be included in multivariate analysis.

But asking multiple questions about the same constructs produces problems such as contextual bias. In the past researchers attempted to eliminate such biases by randomizing the order of the indicators/items, but this approach still makes unreasonable assumptions like the measurement error is the same for each item in each context. This study proposes a new approach to overcome the cited problem. The proposed method calls for asking only one randomly selected attitude question per respondent. Since the missing data is missing by design it is missing completely at random (MCAR). MCAR missing data techniques are used to control for all questions not asked. Monte Carlo simulations show that the approach produces approximately unbiased estimates without substantial loss of power. Results also show that Bayesian estimation of these models is more effective than full information maximum likelihood estimation due to the high degree of failed model convergence with the high levels of missing data.

Effect of Personal Experience on Attitudes: a Path Analysis with Panel Data

Aysel Tutkun; Swiss Federal Institute of Technology Zurich (ETH) , Switzerland

Conversion to particularly animal-friendly stabling (PAFS) and conversion to regularly outdoor exercising of live stock (ROE) are two main animal-friendly programs for which Swiss farmers receive direct payments from the government. These conversions are analysed within a path analysis framework in order to 1) review the link between intention to convert and observed conversion behavior, 2) prove the influence of personal experience on changing attitudes, after having converted.

The main theoretical approach which is used in this study is the Theory of Reasoned Action (ToRA) (Fishbein, Ajzen 1975). This theory is a general theory concerning every kind of social behavior (Ajzen 1980). One of the most important assumptions within ToRA is that people can change their attitudes towards a certain subject after having personal experience with that subject. Roger (1995, 2003) also emphasizes the importance of this assumption within the Diffusion Theory.

In fact, to examine this assumption panel data is required for at least two time points. Therefore two surveys are conducted, first in May 2003 and second in November 2004. In addition observed conversion data from December 2003 as well as from December 2005 is provided from the agricultural ministry of Obwalden so that it is possible to include variables before and after having converted. Swiss farmers can convert to PAFS and to ROE step by step, e.g. first for their cattle, then for their pork and lastly for their poultry. So with this data set one can prove, if farmers did the second conversion step after having converted for the first time, i.e. after having personal experience with PAFS and ROE.

The behavior model encompasses three theoretical constructs which influence the Behaviour: Intention to perform the given behavior, Attitude towards the behavior and Subjective Norm regarding the behavior, which both affect Intention. Lastly, Intention influences the Behavior, i.e. observed conversion to PAFS or ROE. These variables are available for two time points, so one can differentiate between Behavior1 and Behavior2 and so on. In addition, Behavior1 influences Attitude 2 and Subjective Norm 2 of the second time point.

The results indicate a moderate effect of Behavior 1 on Attitude and Subjective Norm which is significant at the 10%-significance level. However, it is also obvious that the stability of the variables between the two time points is quite less maybe they were not well formed at that time and were in the process of forming.

Issue Ownership and Stability and Change in Party Choice

Sarah Nicolet; University of Geneva, Switzerland

Pascal Sciarini; University of Geneva, Switzerland

This contribution studies the extent to which the “issue ownership” model explains patterns of stability and change in party choice across elections. According to the model, voters are likely to vote for the party that is perceived to be better than its opponents at handling the problems facing the country (Petrocik 1996). As a consequence, changes in the electorate’s issue concerns between two elections are expected to produce changes in party choice; parties that “own” the issues that have gained prominence on the political agenda between two elections are likely to attract new voters while parties that are considered competent at handling issues of decreasing importance are likely to lose voters. However, not all citizens are equally susceptible to be affected by issue ownership effects. First, party identifiers are less likely to be swayed by changes in issue prominence; their identification is expected to push them to remain loyal to their party despite changes in issue salience. Secondly, political sophistication is likely to mediate the impact of issue ownership; politically sophisticated voters are likely to be more aware of parties’ issue positions and thus better able to account for issue ownership in their vote. Our hypotheses are tested on the 2003 Swiss elections study (Selects) panel data, in which citizens surveyed during the 1999 federal elections were interviewed again after the 2003 elections. Using multivariate statistical techniques, we examine the impact of issue salience and ownership on stability and change in party preferences between the 1999 and 2003 elections.

Reference

Petrocik, John R. 1996. Issue ownership and presidential elections, with a 1980 case study. *American Journal of Political Science* 40 (3):825-850.

8.5 Time Use Surveys

Session coordinated by:

Boris Kragelj; University of Ljubljana, Slovenia

Mario Cools; University of Hasselt, Belgium

Presentations

1. *Introduction into the Session* – B. Kragelj (Slovenia)
2. *Increasing the Effectiveness of Time-use Survey with Qualitative Methods* – G. Mugan (Turkey), F. Erkip (Turkey)
3. *Different Modes of Time Use Data Collection: a Comparison of Explanatory Potential for Family Related Outcomes* – U. Krieger (Germany)
4. *Electronic Methods for Collecting Time Use Data - a Comparison of Day Reconstruction and Experience Sampling Methods* – S. Mceachern (Australia), D. Lynch (Australia)
5. *An Activity-based Approach for Surveying and Modelling Travel Behaviour: Analysis of the Mixed-mode Design-effects* – M. Cools (Belgium), E. Moons (Belgium), G. Wets (Belgium)
6. *Inserting Very Short Trips into Existing Time-use Surveys* – S. Koelet (Belgium), I. Glorieux (Belgium)
7. *NOMAD: an Integrated Spatial-temporal Activity-based Data Collection* – J. Spinney (Canada)
8. *Kronos and Metis, Two Web Applications for Measuring Study Time* – H. Schrooten (Belgium), K. Vangeel (Belgium)

Presentations in the session "Time Use Surveys"

Introduction into the Session

Boris Kragelj; University of Ljubljana, Slovenia

Time use data is playing an important role in sociological research (studies on the relation family - work, gender equalities, leisure time, quality of life, media studies.), as well as in the field of economics (working hours, paid vs. unpaid work, household production.), urban planning (traveling), social care (young and old people's care) and the related policy issues. Time use surveys are aimed at collecting the data on how people spend their time to produce reliable information for the above research fields.

There are three different types of time use surveys - three different ways to collect time use data: (1) standard survey questionnaires or stylized estimates, (2) time diary surveys and (3) activity logs surveys based on the use of mobile electronic devices.

(1) In standard survey questions respondents are asked how much time they have spent for a particular activity on the previous day or week; or how much time they normally spend each day or week. This approach has higher response rate than others but certain recall biases because it is difficult for a respondent to remember and report exactly the amount of time used for certain activity in a day or week. (2) Time diary surveys asks respondents to write a full time diary of all daily activities or fill in activities and time used of the resent/past day in time slots of a diary. Compared with the stylized estimates this approach is less dependent on the respondents recall and calculations of time spent on different activities and usually it produces more accurate measures. Also, it provides detailed information of the respondent's use of time on all various activities during the day (not only the one asked by survey questionnaire), as well as the sequence of these activities and the contexts of performing them (e.g., with whom and where time was spent). On the other side completing a diary for a whole day or even a week can be too big burden for respondent, resulting in low response rates. (3) Most recent development in the field of time use surveys, activity logs, is precise, puts less burden on the respondents, but they are limited in power as they usually focus only on limited number of activities in a certain period of the day. With this approach we cannot expect respondents to carry the electronic devices everywhere, all the time, and even record their private life.

Taking into account special characteristics of each approach, it is an important issue to find out, if they produce the same results? What are the advantages and disadvantages of each survey method comparing to the others? Considering diary approach as the most reliable -can we estimate measurement errors of others? Can we propose some correction factors that would make (more flexible) stylized estimates more reliable? How can different methods be combined to get the best outcome?

Increasing the Effectiveness of Time-use Survey with Qualitative Methods

Güliz Mugan; Bilkent University, Turkey

Feyzan Erkip; Bilkent University, Turkey

The combination of multiple research methods through employing a hybrid with different aspects of each has a significant role in strengthening a study and compensating limitations and faults. Data generated by qualitative and quantitative methods that focus on the same research problem are the means of enhancing the validity of the results. The process of 'triangulation' refers to this issue of consistency of evidence (i.e. validity). Triangulation is an important reason to combine qualitative and quantitative methods to demonstrate convergence in results. In addition, this combination of methods in a single study has the purposes to use methods sequentially, to find contradictions and fresh perspectives and to add breadth to a study.

'Time-use surveys' that have been carried out in many countries combine various qualitative and quantitative methodological perspectives in data collection, sampling procedures, coding and analysis of data. Although, there is an ongoing debate on appropriate data collection methods for time-use surveys, time diaries seem to be the most common way of collecting this kind of records.

Time diary surveys usually involve combination of qualitative and quantitative methods at different steps of the research process. In diaries, in-depth qualitative data such as daily life stories of the respondents are expected in the form of recording their course of the day through subjective verbal descriptions. These diaries are accompanied by questionnaires (regarding household structure and members), observations and/or pictograms. Depending on the expected representativeness of the sample and the scale of the study, either quantitative or qualitative sampling methods can be applied. These flexible and mixed methods of data collection and sampling are, however, associated with a complex quantitative modelling of data interpretation and analyses. This research focuses on the analysis of potential problems and additional issues with qualitative methods to enhance the first national time-use survey that is in progress in Turkey. The Turkish Statistical Institute (TURKSTAT) initiated the first national time-use survey in December 2005 and completed the field survey recently. It consists of a 24-hour diary with follow-up interviews with 5070 households. Although it appears to be a mature application, there are contextual problems in the design and the application of it. This paper aims to overview how this survey can be enhanced with qualitative methods at the Turkish cultural context. Therefore, we have conducted a pilot study with a design of 'dominant-less-dominant' in which dominant qualitative paradigm was considered including its sampling and data collection methods. Quota and convenience sampling methods were used for approximately 70 individuals. For data collection methods of the study, activity listing which is followed up by face-to-face in-depth interviews was accompanied by quantitative household and individual questionnaires.

By the field survey, we hope to contribute to improve the national time-use survey in collaboration with TURKSTAT officials. While testing the alternative methods concerning the time-use of variety of household members from different areas of Ankara, the capital city of Turkey, we try to give special emphasis on one of the most ignored component of time-use studies; namely data concerning space-use other than geographical location. In all these respects, what expected is to demonstrate the methodological triangulation that occurs between qualitative and quantitative approaches in the way of sequential triangulation. As a result, we aim to argue that the effectiveness of time-use surveys can be increased with qualitative interventions.

Different Modes of Time Use Data Collection: a Comparison of Explanatory Potential for Family Related Outcomes

Ulrich Krieger; University of Mannheim, Germany

Information on time use is essential for understanding many processes in families and durable relationships. However, studies differ substantially in how they collect information on time use. We use newly available data from the German Mini Panel study on families and relationships, (a preparatory three wave panel for the planned German Family panel). Given the experimental character of the study, both standard survey questions on time use and an elaborated time use diary have been administered to respondents and their partners as part of this study. Furthermore, the Mini Panel contains information on important family outcomes such as relationship forming, stability and fertility.

As in similar research, substantial differences in answers on time use of the individuals can be shown between the diary and survey question answers. But rather than taking the difference in answers as dependent variable and explaining the difference, this study has a different focus.

Using the two sets of information on time use (diary and survey question) we examine which set of time use information is better suited as independent variable explaining outcomes relevant to family research, namely marital satisfaction and fertility. Panel Regression models on we want to discriminate which time use information better helps us to understand these outcomes, making full use of the longitudinal nature of the data. Possible implications for further time use research are being discussed.

Electronic Methods for Collecting Time Use Data - a Comparison of Day Reconstruction and Experience Sampling Methods

Steven Mceachern; University of Ballarat, Australia

David Lynch; University of Ballarat, Australia

The increasing use of time use data in sociological and economic research has made the accurate collection of such data an increasing priority. The challenges of collecting time use data are relatively well understood in the literature, with the dominant methods - questionnaires and diaries - each exhibiting varying levels of data quality.

A promising development however for time use research is the continuing improvement in electronic data collection methodologies. Questionnaires administered by email, the web and mobile telephones or PDAs provide new alternatives for time use researchers, but bring with them another set of challenges.

This paper reports on a pilot study of time use over a two week period among 70 university students in regional Australia. The study utilised two alternate forms of time use data collection, firstly using an online diary study following the day reconstruction method (Kahneman et al, 2004), and then via personal digital assistant (PDA) using the experience sampling methodology (Larson and Csikzentmihalyi, 1983). Each student completed two weeks of time use data collection, one week using each method.

Analyses in the paper compare differences both between and within respondents, in terms of reporting practices and response rates, using the two methods. The paper then continues on with an examination of the differential patterns of reported time use between the two methods, and then concludes with a discussion of the relative efficiency and viability of each of the two approaches for time use studies.

References

- Kahneman, Daniel, Alan B. Krueger, David A. Schkade, Norbert Schwarz, and Arthur A. Stone (2004) A Survey Method for Characterizing Daily Life Experience: The Day Reconstruction Method. *Science*, 306, 1776-1780
- Larson, R. & Csikzentmihalyi, M. (1983). The experience-sampling method. In H. Reis (Ed.), *New directions for naturalistic methods in the behavioral sciences*. San Francisco: Jossey-Bass.

An Activity-based Approach for Surveying and Modelling Travel Behaviour: Analysis of the Mixed-mode Design-effects

Mario Cools; University of Hasselt, Belgium

Elke Moons; University of Hasselt, Belgium

Geert Wets; University of Hasselt, Belgium

Reports from various international organisations, like for instance the European Commission's White paper "European transport policy for 2010: time to decide" show that policy makers acknowledge the increasing importance of mobility. Mobility is not just considered to be a cornerstone for economic growth, but also seen as a social need that offers people the opportunity for self-fulfilment and relaxation.

In order to lead an efficient policy, governments require reliable predictions of travel behaviour, traffic performance, and traffic safety. Thus, traffic and transportation models can provide the right framework to support long-term decisions. On an international level, activity-based models have become one of the leading paradigms to model travel behaviour. The most important characteristic of these models is that travel is considered as a derivative from the activities that individuals and households need or wish to perform. This means that travel is no longer seen as an isolated fact in these models, which is a great advantage in comparison to the classic models.

The development of activity-based models requires very specific data. Since a special emphasis was laid on the dynamic character of the model, both the planned and executed activities were surveyed. 2500 households were selected using a stratified cluster technique. The selected household were asked to fill out an activity-diary and to report rescheduling decisions during a one-week period. Approximately one half of the households received a PDA-module; the other half was surveyed by means of a traditional paper-and-pencil diary.

Although both data collection modes were designed in such way that the correspondence of the questions was optimised, it remains essential to identify differences in the data that are introduced by using multiple modes. This paper starts with describing the differences that existed in the questionnaire design of the two modes, and then further explores the statistical differences in terms of some key observations, such as average number of trips reported, and average travel time.

Inserting Very Short Trips into Existing Time-use Surveys

Suzana Koelet; Vrije Universiteit Brussel, Belgium

Ignace Glorieux; Vrije Universiteit Brussel, Belgium

Transport researchers increasingly acknowledge the advantages of activity-based diaries to collect travel information (Stopher 1992) and to model travel patterns (Lawton & Pass 1997). This type of information is readily available to researchers thanks to large-scale national time use surveys administered in a large range of European and non-European countries (Fisher 2004). Although not conceived for this type of research, the detail of the activity context in time use surveys is very useful for transportation modelling and makes the travel data less prone to memory flaws. But because they are not conceived for transportation research, time use surveys can also have important drawbacks for the measurement of travel. In this paper we focus on the possibility given to respondents in some time use surveys not to register activities shorter than 5 minutes. Short trips are nevertheless of great importance to transport researchers. They represent a large share of total mobility and their impact on basic mobility indicators is considerable (Armoogum, Castaigne, Hubert 2005; BGC 1995). As such they are a key variable in research on traffic congestion, environmental sustainability or mobility. The contextual information available in time use surveys however offers us the opportunity to recover short trips missed by this recommendation. Short trips are added to the 1999 Flemish time-use survey every time a respondent changes location and mentions a transportation mode, without writing down an actual transportation activity. This increases the amount of trips in the survey with no less than 27%. For no less than 76% of the respondents at least one short trip is added during the registration week by this transformation. The impact of this major intervention is measured on some basic mobility indicators. The results are compared with respectively data from a time-use survey administered in the same region that year, that does not use the '5 minute'-recommendation and data from a travel survey in the same region a year later. We also try to predict immobility before and after adding the short travel activities to the survey to get an indication of the validity of this transformation. This will also gives us more information on the incidence of short trips in different socio-economic groups.

NOMAD: an Integrated Spatial-temporal Activity-based Data Collection

Jamie Spinney; Saint Mary's University, Canada

Time use surveys are regularly conducted by national statistics agencies and concentrate on the timing of the daily stream of activities using time diaries. Travel surveys are regularly conducted by municipal planning agencies and concentrate on the location of trip origins and destinations using travel diaries. Meanwhile, spatial-temporal activity surveys are rarely conducted, even though they are essential for informed land use planning. Over the past decade, metropolitan planning strategies are increasingly considering combinations of land use and transportation policies, which has lead to a growing interest in a spatial-temporal activity-based data collection framework. Furthermore, recent advances in computer assisted telephone interviewing (CATI), geographic information systems (GIS), global positioning systems (GPS), and cellular technologies offer considerable advantages for collecting spatial-temporal activity data. This presentation reports on an innovative prompted-recall approach to capture geo-referenced time-use data that is able to satisfy time use research, and both land use and transportation planning applications. By integrating the data collection, data entry, activity coding, geo-coding, and data visualization into a single CATI software package called NOMAD©, we have developed an innovative prompted-recall approach to capture geo-referenced time-use data. For example, our customized

GPS takes advantage of cellular technology to transmit the GPS data so they can be pre-processed to identify stops (i.e. destinations) and travel routes, which are then displayed within the NOMAD system to enable prompted-recall of the primary respondents' detailed activity patterns. This innovative prompted-recall approach is examined insofar as its contribution to improving data quality and reducing respondent burden.

Kronos and Metis, Two Web Applications for Measuring Study Time

Hilde Schrooten; University of Leuven, Belgium

Koen Vangeel; University of Leuven, Belgium

How much time is studying that course going to take? Students often wonder when they collect their study materials. The institution and the lecturers are also interested. They have carefully planned the course units, taking into account the time that students should spend on each course unit. This time to be spent is expressed in credits. One can expect that for each credit students spend 25 to 30 hours on average in the course of the academic year. A Flemish decree ordains that a regular study programme consists of 60 credits (1500-1800 hours/year). Lecturers want to know whether their estimation of the necessary time corresponds with the actual time spent. Flemish institutions are also obliged by decree to examine this correspondence. Therefore the institutions conduct study time measurements with the students.

Recent developments have reinforced the necessity to collect objective data on study time. More and more students are compiling individual study programmes in a flexible way. The institution has to guarantee that the study load of a programme that consists of 60 study points is feasible, irrespective of which course units the students combine. Moreover, the Bologna agreement aims at stimulating student mobility in Europe. When students get credits in other European countries, it is important that the necessary study time corresponds with the credits of the course units.

In a project on "Study time and feasibility of study load" of the K.U.Leuven Association 12 university colleges and the University of Leuven joined forces to develop two web applications for measuring students' study time. A first application, Kronos, is based on the technique of weekly time registration. Students are asked to take down their study time for the courses they attend. Mostly they are asked to classify the study time in three categories: study time while attending the lecturer's sessions, work on compulsory assignments outside the student-lecturer contact time and individual processing of the subject matter.

A second system is called Metis, the system that supports three time estimation techniques. The first one is an absolute estimation technique. At the end of a certain period of time (a module, a semester, an academic year) students are asked to assess the time they spent on each course unit in their programmes. To enhance the reliability of the assessment, lecturers provide a detailed list of educational activities in their course within the same three categories. Students estimate the time spent for each subcategory. They are also provided with suggestions on how to estimate the study time, e.g. 'subtract the number of hours you were absent (due to illness, skipping classes, etc.) from the total contact time'.

Metis also offers two relative assessment techniques: the technique of paired comparison and the sorting data technique. In the paired comparison technique students are asked to compare all course units within their study programme two by two and select for each pair the course unit on which they have spent most study time. In the sorting data technique students are asked to sort all course units they attended according to the amount of study time spent. These relative assessment techniques are mostly combined with the absolute estimation of the time spent on some course units that are used as a gauge. Afterwards, the relative scale can be transmitted into absolute study times using techniques of curve fitting.

The web applications minimise respondent burden. Firstly, the student interface is user friendly: information and help is provided, students can deliver the necessary information by a minimum of mouse clicks, if necessary they can stop their registration session and finalise it some other time. Secondly, an English version is available for non-Dutch speaking students. Thirdly, Kronos and Metis can be used

at home or on campus (individually or in organised group sessions). The systems can be used not only by students who regularly attend classes at campus, but also by fellow students in an international exchange project or last-year students that follow a practical training in an organisation. Fourthly, Kronos supports sampling techniques in order to avoid respondent fatigue and diminishing response.

The systems also minimise research time and, in doing so, reduce research cost. Firstly, they automatically generate response tables that the researcher can use to stimulate the response: the list indicates which students participated (and for Kronos also in which weeks they participated). Secondly, the researcher can use the mail function to send thank you and reminder mails. Thirdly, the systems reduce the administration of student data (drop outs and new registrations), course data (followed courses and exemptions), samples (which students are in which sample) and credits (which students have credits for which course units). Last but not least, the data analysis is automated. The systems process the data of the successful students (or other groups of students, e.g. non-successful students, students with deficiencies in their previous education) and generate average study times and other measures of the central tendency and deviation of study times. Moreover they generate graphs comparing the estimated study time based on credits with the subsequently estimated study time actually spent, and for Kronos also graphs on the evolution of the study time. Reliability measures are also included.

Kronos and Metis have been tested in nearly 300 pilot studies with more than

1. researchers. Researchers have convened in user groups and have provided

feedback on the system, which has led to numerous adjustments. User satisfaction is very high. Results on response, average study times and reliability measures are currently being collected and analysed. In the presentation the system will be demonstrated and results will be discussed.

8.6 Measuring and Explaining the Fear of Crime

Session coordinated by:

Jonathan P. Jackson; London School of Economics, United Kingdom

Stephen Farrall; University of Keele, United Kingdom

Emily Gray; University of Keele, United Kingdom

Presentations

1. *Scaling Fear: Overcoming Data Complications in the Development of Indices* – J. Kuha (United Kingdom)
2. *Correlates of Fear: Rounding Up the Evidence from 25 Years of the British Crime Survey* – J. Jackson (United Kingdom), K. Hohl (United Kingdom)
3. *New and Old Measures of the Fear of Crime: a Multilevel Comparison of Measures of Intensity and Frequency* – I. Brunton-Smith (United Kingdom)
4. *Experience and Expression in the Fear of Crime: on the Social and Cultural Meaning of the Fear of Crime* – E. Gray (United Kingdom), J. Jackson (United Kingdom), S. Farrall (United Kingdom)

Presentations in the session "Measuring and Explaining the Fear of Crime"

Scaling Fear: Overcoming Data Complications in the Development of Indices

Jouni Kuha; London School of Economics, United Kingdom

The British Crime Survey includes eight questions on respondents' levels of worry about eight specific types of crime. These can be used to derive one or more overall scores of fear of crime, for example by using methods of latent-variable modelling. However, the data present a number of complications which need to be addressed in order to take full advantage of the technical strengths of the indicators. For example, several of the survey items (e.g. on car crime, racially motivated attack, and rape) may have rather different meanings for different groups, or even no relevance at all for some of them. This paper first shows that such differences exist in the British data, and second describes appropriate methods of analysis. Finally, the use of the measures derived from the analysis is illustrated by an examination of how fear of crime varies by age and gender in the British population.

Correlates of Fear: Rounding Up the Evidence from 25 Years of the British Crime Survey

Jonathan P. Jackson; London School of Economics, United Kingdom

Katrin Hohl; London School of Economics, United Kingdom

Since the beginning of the 1980s the British Crime Survey (BCS) has tracked trends over time in fear of crime. Analysts have highlighted which social groups report more or less fear or have more or less confidence. Such information plays an important role in criminal justice policy. Yet BCS data go somewhat under-utilised by the academic community. As is so often the case with publicly funded social surveys, few researchers analyse the BCS, meaning less effort is put in to analysing the data as has been put into collecting the data. Particularly striking is the lack of definitive statements on key correlates of the fear of crime. For example, evidence is mixed on the relationship between age and fear, with the results seemingly dependent on the measures used, whether bivariate or multivariate analysis is conducted, and whether non-linear and interaction effects.

This paper rounds up the evidence from all available sweeps of the BCS on key correlates of worry about crime (measured via standard indicators). Two innovations are the assessment of changes in the correlates over time and the pooling of data to conduct a meta-analysis.

New and Old Measures of the Fear of Crime: a Multilevel Comparison of Measures of Intensity and Frequency

Ian Brunton-Smith ; University of Surrey, United Kingdom

In a recent sweep of the British Crime Survey (BCS), new questions designed to capture the 'episodic' nature of fear of crime were fielded, examining both the frequency and intensity of specific incidents of worry. These highlighted some interesting contrasts with the existing intensity of worry questions that have been increasingly used to measure fear, identifying a significant proportion of respondents that indicate high overall levels of worry about crime but who do not report worrying about crime on any specific occasions. This suggests that the new measures may be tapping into a somewhat different element of fear, with a separate set of influences that can provide additional detail about the complex nature of fear of crime.

To help understand the differences in the type of fear that are captured using the two measurement strategies, the current paper presents a direct comparison of the question types. BCS data is used to highlight the differences in the individual characteristics that are correlated with the standard intensity of worry questions, and those correlated with the items measuring the frequency and intensity of specific incidents of worry. In addition to identifying differences in the individual correlates between the question types, the current analysis demonstrates the important and differential impact that local context has,

using multilevel models to highlight the relative influence of area differences on each fear measure. Area characteristics drawn from census data and other administrative sources are then used to identify further important contrasts between the question types.

Experience and Expression in the Fear of Crime: on the Social and Cultural Meaning of the Fear of Crime

Emily Gray; University of Keele, United Kingdom

Jonathan P. Jackson; London School of Economics, United Kingdom

Stephen Farrall; University of Keele, United Kingdom

This paper begins by distinguishing between two manifestations of the fear of crime; one is intimately grounded in the daily experience of everyday life, the other is better described as more diffuse anxiety or expressive fears regarding the cultural meaning of crime. Drawing on data from the 2003/2004 British Crime Survey, we test a framework that models both types of 'fear' with objective and subjective assessments of the environment. Using structural equation modelling, we explore the extent to which levels of crime, deprivation and long-term changes in the neighbourhood shape individual assessments of disorder and social cohesion. Second, 'events' of disorder are expected to signal low levels of cohesion. Thirdly, we predict individuals' perceptions of the likelihood of victimization are shaped by their own interpretations of disorder and cohesion in their community. Fourth, perceived risk and assessments of the environment are expected to predict the two routes of fear of crime: everyday events of worry and a more generalised social anxiety. Finally, victimisation and knowledge of local victims is expected to predict the everyday experiential worry, but not the more expressive dimensions of fear. The paper concludes with the argument that the fear of crime – whether it is a routine experience (i.e. remembering to lock the house) or a more communicative anxiety (i.e. deteriorating social bonds) – expresses a set of lay diagnoses of social order and control. However, the more experiential aspects are more highly related to actual experiences of victimisation (and to knowing victims).

8.7 Survey Data and Public Dissemination

Session coordinated by:

Paula Devine; Queen's University Belfast, United Kingdom

Presentations

1. *Disseminating Survey Data for Public and Policy Audiences* – P. Devine (United Kingdom)
2. *Building a Web Resource for Social Scientists: Displaying Survey Questionnaires to Data Users* – J. Lamb (United Kingdom), G. Hughes (United Kingdom)
3. *Record-linkage for Assessing Re-identification Risk in Business-surveys* – T. Bachteler (Germany), A. Dundler (Germany), D. Engel (Germany), J. Reiher (Germany), R. Schnell (Germany)
4. *A Questionnaire Development Documentation System* – R. Schnell (Germany), A. Zwingenberger (Germany), O. Hopt (Germany), M. Stempfhuber (Germany), J. Krause (Germany)
5. *Quality Assessment of Statistical Information Through the View of Its Users'* – A. Stoykova-Kandlieva (Bulgaria)

Presentations in the session "Survey Data and Public Dissemination"

Disseminating Survey Data for Public and Policy Audiences

Paula Devine; Queen's University Belfast, United Kingdom

The aim of ARK (Northern Ireland Social and Political Archive, www.ark.ac.uk) is to make social and political material on Northern Ireland available to the widest possible audience. Our audience includes academic and government policy makers and the general lay public, which includes constituencies such as school children, the media and local voluntary groups. Thus information needs to be available in a variety of appropriate formats. One key activity of ARK is running and disseminating the results of the Northern Ireland Life and Times (NILT) Survey, which is an annual survey recording the attitudes of the public of Northern Ireland to a range of social policy issues.

Using the example of the NILT survey, this paper will start with the premise that good research leads to public debate which leads to policy development which in turn leads to further good research. The paper will identify the various methods by which ARK has disseminated results from the survey, and the rationale behind each one. In particular, the paper will highlight how the government and voluntary sectors have made use of the results to develop, test awareness of and evaluate a range of policies. The paper will also explore how statistical information can be made accessible for a wider lay audience, and why it is important to do so.

Building a Web Resource for Social Scientists: Displaying Survey Questionnaires to Data Users

Julie Lamb; University of Surrey, United Kingdom

Graham Hughes; University of Surrey, United Kingdom

The ESRC Question Bank (Qb) is a web based survey resource which exists to make survey questionnaires and other key metadata easily available to survey researchers. The resource is free to use with no registration. This paper will provide a demonstration of the Qb, with a particular emphasis on how the project has developed the user interface of the web site in order to best disseminate the vast amount of information that is currently available. In the past year the Qb has undergone extensive redesign with user feedback in mind.

This paper will have two key aims: firstly it will examine what the Qb is and how it is used in conjunction with other existing data resources; secondly to explain how such a web resource is built up paying particular attention to the documentation of survey interviews, almost all of which are carried out using Computer Assisted Personal Interviewing. It is apparent that there is no standard method of documenting a Computer Aided Interviewing (CAI) program and so there is no straightforward way of reading that documentation in order to identify the circumstances in which specific data items were collected. This is an issue with which the Qb team has been grappling and our findings which have implications for the standardisation of survey documentation will be discussed here.

As the next generation of students expects ever more complex web resources to be available are we as a data community ready for the challenge; how will we be able to best represent CAPI questionnaires? This will be explored and discussed.

Record-linkage for Assessing Re-identification Risk in Business-surveys

Tobias Bachteler; University of Konstanz, Germany

Agnes Dundler ; Institute for Employment Research, Germany

Dirk Engel; University of Konstanz, Germany

Jörg Reiher; University of Konstanz, Germany

Rainer Schnell; University of Konstanz, Germany

Usually, business-surveys contain many factual variables on establishments, which are also available in commercial available data bases. The largest of this data bases is AMADEUS (Analyze MAJOR Databases from EUROpean Sources). AMADEUS contain information on eight million enterprises in 38 european countries; 712.000 of them in Germany (about 80% of all enterprises). In order to estimate the risk of re-identification in business-surveys, we used the largest german business survey as demonstration. This so called "IAB-Establishment-Survey" is a panel study of about 15.000 enterprises in Germany. Using variables like number of employees, regional identifiers, occupational sector and organizational details, a probabilistic record linkage has been attempted. We will report on the conditional probabilities of a correct re-identification given public available information, depending on size and type of enterprises.

A Questionnaire Development Documentation System

Rainer Schnell; University of Konstanz, Germany

Anja Zwingenberger; University of Konstanz, Germany

Oliver Hopt; Social Science Information Centre (IZ), Germany

Max Stempfhuber; Social Science Information Centre (IZ), Germany

Jürgen Krause; Social Science Information Centre (IZ), Germany

Due to pretest results, a questionnaire is usually revised several times. Nearly always, the changes are done within a word processor. Previous versions are rarely preserved and the causes of a change rarely documented. Such documentation would be useful for further research in questionnaire design and also for teaching. In order to allow such extensive documentation of changes, special software is needed. Up to now, such software did not exist. Therefore, the authors developed a special database system (QDDS) for this purpose. QDDS is a highly portable JAVA-program suitable for the documentation of paper&pencil, face-to-face and CATI instruments. A database for each questionnaire contains information on question content, layout and formal aspects of each question, such as number and kind of response categories. Furthermore, it may contain all previous versions and every documented cause for a change. The application of this system to a large scale survey will be presented. Finally, the currently undergoing extension of the system in order to document all instruments within a data archive or research institute will be described. This included the import and export of questionnaires in several file formats, for example different CATI-systems.

Quality Assessment of Statistical Information Through the View of Its Users'

Andreana Stoykova-Kandlieva; University of National and World Economy, Bulgaria

The insurance of the required quality of the produced statistical information is a fundamental matter of this production for every informational system. The attentive analysis of the quality informational matter shows that it has three aspects: 1. the quality of the statistical informational product; 2. the quality of the activity regarding this product's making. 3. the quality of the provided statistical date to the users (quality of user servicing). The assessment of these three aspects of quality can be accomplished through different prisms: one of them is through the users' satisfaction of the statistical informational quality. From this kind of view, the problems, the selection of the units and some more important conclusions from that kind of survey is dedicated to the present report.

In it are exposed some basic formulations and conclusions of the carried out in 2005 by a group of experts in Bulgaria survey of the users' satisfaction from the statistical data quality as carrying out of the project "Regional Statistics Improvement and Assessment of the Quality of the Statistical Date", EuropeAid/116362/D/SV/BG. According to the conclusion of the subjective opinions, assessments and recommendations of 1502 users of the public authorities (ministries, agencies, district authorities and other institutions), local authorities (municipalities) , business enterprises and other institutions (scientific, trade-union, branch, media) are gained assessments for the produced informational quality

of the National Statistical Institute (NSI) in Bulgaria. Within the sample of 1502 users of statistical information, have been determined sub-samples included in the research in a different way. In business enterprises, the survey is two-stage cluster sample, for the central public administration and district administration is exhaustive and for the rest units also as a non-random pilot sample.

Within the establishing of the satisfaction level of the statistical informational quality that NSI offers to its users is being made an attempt for an assessment of the common quality level of the produced statistical data.

As a result of the survey may be drawn out three main groups of conclusion.

The first group is connected with the production's technology of similar kind of survey that concerns definite purposes, assessments and approaches in it. The analysis shows they can also be considered as absolutely satisfactory. The scope of the survey can be considered as a fully satisfactorily as well as its possibilities for separating the customers groups as of the point of view of achievement the desired representativeness of the results. The outlined four users' groups appear as a good base for further profound analysis.

According to the adopted methodology, the examination marks of the components of the statistical data quality may be considered as a good attempt for real estimation of the quality level of the produced statistical data.

The second group concerns some specific results of the survey. As such can be noted the competence and clear requirements of the users' majority of the statistical information on one hand, as the presence of a special users' group on the other, which are in process of forming their statistical culture defining their necessities and requirements towards the given statistical information. In this group users must be directed in the efforts and attention of the satisfactorial informational producers.

As an important conclusion is considered the fact that when assessing the importance of different components in statistical informational quality, the users' proceed moreover from their impressions concerning the real state of quality than the presumable level where the given component should be put in order.

The users' satisfaction of the given statistical information is assessed through two prisms: users' satisfaction of the informational service which NSI offers and the assessment of definite statements connected with the quality of statistical data.

The third group of conclusions are connected with the future work of the NSI for improving the quality of the produced and provided statistical data. Some basic directions are outlined, where users might want to change the produced and provided information from NSI. Basically these are changes in the time limit for preparation and presentation of the statistical data followed by the precision and reliability of the information, the range of the used indexes and the comparing of the statistical data.

In the same time the carried out research could be repeated in a definite period of time, for example 2 years. In this way it could be traced how NSI reacts on the users' necessities and could be established how effective are the change of this institution concerning the statistical informational quality.

Keywords: quality of the statistical information; components of the quality of the statistical data; assessment of the quality of the statistical information, users of statistical information; users' satisfaction with the quality of statistical data.

8.8 Survey Evaluation Methods Using Behaviour Coding and Sequence Analysis

Session coordinated by:

Emanuela Sala; University of Essex, United Kingdom

Noah Uhrig; University of Essex, United Kingdom

Presentations

1. *The Development and the Implementation of a Coding Scheme for Behavioural Coding, the Experience of the British Household Panel Survey* – E. Sala (United Kingdom), N. Uhrig (United Kingdom)
2. *Combining Computer Assisted Recorded Interviewing (cari) and Behaviour Coding to Measure Data Quality: a Discussion of the Benefits and Lessons Learnt* – A. McGee (United Kingdom)
3. *Verbal Behavior Coding when Data Are Collected with Event History Calendar* – M. Callegaro (United States), R. Belli (United States)

Presentations in the session "Survey Evaluation Methods Using Behaviour Coding and Sequence Analysis"

The Development and the Implementation of a Coding Scheme for Behavioural Coding. the Experience of the British Household Panel Survey

Emanuela Sala; University of Essex, United Kingdom

Noah Uhrig; University of Essex, United Kingdom

Behaviour coding is a technique of observing and recording action, both verbal and non-verbal. The technique utilises discrete taxonomies to record such action thus making it amenable for systematic analysis. Behaviour coding has been widely used to monitor and evaluate survey interviewing as well as to pre-test questions for several decades. Despite its long-standing use, little is written about the procedures to be followed while developing a coding scheme. A notable exception is Cannell et al.'s classic work introducing their method, however this is over 30 years old.

Our paper aims to provide a practical background on the development and the implementation of the behaviour coding scheme. The coding scheme was developed to explore interview dynamics and, in particular, to analyse the sequence of the question-answer process and the different ways in which such order differs from the so-called paradigmatic sequences. The scheme was used for coding 152 interviews of British Household Panel Study Wave 16 pilot conducted in March 2006. We will briefly describe the overall project that evaluates these interviews. We will then provide a review of several coding schemes that have been used in the past. We will describe the procedures that were followed while developing the coding scheme to analyse the BHPS transcripts and illustrates the procedures adopted to train the coders. We conclude with a discussion on the lesson learnt.

Combining Computer Assisted Recorded Interviewing (cari) and Behaviour Coding to Measure Data Quality: a Discussion of the Benefits and Lessons Learnt

Alice McGee; National Centre for Social Research - NatCen, United Kingdom

Measuring the quality of survey data is an activity that is increasingly becoming a standard component of the survey research process. However, how to define and measure survey quality, or aspects of it, is complex. This paper focuses on, and evaluates, the combination of two relatively new methodological tools, sometimes used for the purpose of measuring survey interview data quality: Computer Assisted Recorded Interviewing (CARI) and Behaviour Coding.

CARI is a software application that allows the computer to act as a sophisticated tape recorder while the interviewer administers a CAPI (Computer Assisted Personal Interview) questionnaire. CARI unobtrusively records the verbal interaction between the interviewer and respondent during interviews (Herget, Biemer, Morton and Sand, 2001). CARI audio quality can be comparable to using tape recorders, but without the practical problems associated with the use of external audio recording devices in large-scale field operations.

Behaviour coding is used to analyse the interactions between interviewers and respondents during the administration of survey questions (Cannell, Fowler, and Marquis, 1968). A code frame is systematically applied to interviewer and respondent behaviours displayed during the question and answer process (Oksenberg, Cannell, and Kalton, 1991; Sykes and Morton-Williams, 1987).

This paper will present a recent piece of work that combined these two methods to assess the data quality of specific question items, outlining the advantages and disadvantages of the approach. Through a discussion of the methodological and practical challenges of the work, it will highlight the lessons learnt and put forward suggestions for the contribution these methods could make to future work.

Verbal Behavior Coding when Data Are Collected with Event History Calendar

Mario Callegaro; University of Nebraska, Lincoln, United States

Robert F. Belli; University of Nebraska, Lincoln, United States

The event history calendar is a flexible interviewing approach that is designed to collect retrospective reports of events and the timing of their occurrences for reference periods that can range from several months to the entire life course. Key components of event history calendar design include representing thematic aspects of the past into domains, such as residence and labor, and the capturing of temporal changes by the inclusion of one to several timelines within each domain. EHC interviewing style is more towards the flexible approach. Standard behavioral coding schemes do not capture the richness of data coming from an EHC interview.

In this paper the development and application of a EHC behavioral coding scheme is described with results from a comparison of data quality between EHC and conventional questionnaire interviews.

8.9 Difficult Groups in Survey Research

Session coordinated by:

Remco Feskens; University of Utrecht, Netherlands

Hans Schmeets; Statistics Netherlands (CBS), Netherlands

Presentations

1. *Difficult Groups in Survey Research* – R. Feskens (Netherlands), H. Schmeets (Netherlands)
2. *Non-response in a Survey Among Immigrants in Denmark* – M. Deding (Denmark), T. Fridberg (Denmark), V. Jakobsen (Denmark)
3. *Fieldwork Substitution As an Instrument to Cope with Selective Dropout in Surveys* – S. Demarest (Belgium), J. Tafforeau (Belgium), J. Van der Heyden (Belgium), S. Drieskens (Belgium)
4. *Research Among Ethnic Minorities in the Netherlands* – J. Kappelhof (Netherlands)
5. *Surveying Victimization Among Immigrant and Ethnic Minority Groups* – J. Goodey (Austria), S. Nevala (Austria)
6. *Labour Market and Social Security: a New Panel Study for Research on German Social Code II* – R. Schnell (Germany), J. Achatz (Germany), B. Christoph (Germany), M. Trappmann (Germany), C. Wenzig (Germany)
7. *Where Have All the Young People Gone?! an Analysis of the Difficulties in Locating and Gaining Access to the Juvenile Segment in Telephone Surveys.* – S. Pasadas Del Amo (Spain), M. Soria Zambrano (Spain), M. Trujillo (Spain), M. Zarco Uribe-Echevarría (Spain)

Presentations in the session "Difficult Groups in Survey Research"

Difficult Groups in Survey Research

Remco Feskens; University of Utrecht, Netherlands

Hans Schmeets; Statistics Netherlands (CBS), Netherlands

For various reasons, Statistics Netherlands has particular difficulties surveying five subgroups, the so-called difficult groups:

- Ethnic minorities
- People living in institutional households
- The homeless
- Asylum seekers
- Illegal persons

The above-mentioned difficult groups are underrepresented in survey research because of – among other things – nonresponse and undercoverage. Results of the survey objectives will possibly be biased if the response is not equally distributed among various societal groups.

Because of the potential worrying consequence of biased results, Statistics Netherlands started the strategic research project Nonresponse and difficult groups in collaboration with a number of universities. The aim of this project is to enhance the quality of survey estimates in general, and the statistical information about these difficult groups in particular.

In this presentation we shall focus on the possibilities and difficulties of surveying difficult groups. We shall discuss the effect of an increase in the number of contact attempts and the use of incentives on the response rates among ethnic minorities. Furthermore, we shall show how we have developed a guideline to study people living in institutional households, a difficult group which is often excluded from the sample frame before the actual study takes place. Finally we discuss results of the Dutch Parliamentary Election Study 2006, where some of our findings were implemented to enhance the quality of the data.

Non-response in a Survey Among Immigrants in Denmark

Mette Deding; Danish National Institute of Social Research (SFI), Denmark

Torben Fridberg; Danish National Institute of Social Research (SFI), Denmark

Vibeke Jakobsen; Danish National Institute of Social Research (SFI), Denmark

During the last 20 years, the number of immigrants from non-Western countries in Denmark (as in other European countries) has increased significantly. Consequently the need for knowledge about the immigrant groups has also increased. However, the basic lesson learned from surveys conducted among immigrant groups is that non-response rates are typically relatively large among immigrants, and thus the surveys have revealed that interviewing immigrants requires other considerations than interviewing the majority population.

The purpose of the paper is to study non-response in a survey among immigrants in Denmark. Data used for the analysis is survey data from Denmark including immigrants from Turkey, Iran and Pakistan as well as ethnic Danes. The individuals selected for interview are aged 18-45 years. The present survey more than fulfilled the expectation that non-response would be high. The response rate was roughly 50 percent for the immigrants and 80 percent for ethnic Danes. The survey collecting process also revealed that there are large differences in non-response and reasons for non-response for immigrant from different countries and for men and women. The aim of this paper is to analyse these differences and thus enable future surveys to be tailored to the specific group in greater detail.

In the paper, we analyse (using multilevel logit model) how characteristics of the individuals and of the interviewers' are related to response-non-response using register information for all the individuals selected for interviews. To focus on the difference between the different non-response reasons, we estimate two models. The first is the probability of response for the entire sample; while the second is the probability of response given contact has been established.

The primary results from the first model shows that the response rate is significant highest for young people, individual that live in couple and have children, individual with more education, employed individuals and Danish citizen. Individuals living in Copenhagen have a significant lower response rate than individuals living in other urban areas or rural areas. Country dummies show that the response for all three immigrant groups was lower than for Danes, indicating that there is a negative effect on response not being captured by the other variables.

Comparing results from the second model with results from the first model reveal if refusals differ from overall non-response in terms of relevant factors. Comparing the two sets of results, we find that some factors matter for overall response but not for response given contact. These are age, family situation, Danish citizenship and country dummies for Iran and Pakistan. This indicates that these variables are more important for the contacting phase than for the persuasion phase given contact.

The interviewer variables (seniority, age, gender, numbers or interview per interviewer and number of municipalities per interviewer) are not significant when controlling for the characteristics of the respondents.

Fieldwork Substitution As an Instrument to Cope with Selective Dropout in Surveys

Stefaan Demarest; Scientific Institute of Public Health, Belgium

Jean Tafforeau; Scientific Institute of Public Health, Belgium

Johan Van der Heyden; Scientific Institute of Public Health, Belgium

Sabine Drieskens; Scientific Institute of Public Health, Belgium

Fieldwork substitution refers to a survey approach in which non participating units are replaced by new (matching) units (called substitutes). These substitutes are contacted after originally selected units fail to respond. This technique has often been criticized for it does not necessarily eliminate non response bias, it should tend to decrease the effort of interviewers to obtain participating from originally selected units, it is prolonging the fieldwork,... Experience from the Belgian Health Survey (HIS), a large scale survey that uses fieldwork substitution, will enable us to estimate the consequences of using this technique.

In the HIS2004 households were selected as (secondary) sampling units. For each selected households 3 matched substitutes were selected, constituting a cluster. Households from the same clusters had specific common characteristics: they belong to the same statistical sector (neighborhood), have a reference person (head of the household) of the same age group and an identical household size. Interviewers were not aware of the status of households (initially selected household or substitute) when conducting the fieldwork.

In order to evaluate fieldwork substitution, the composition (in terms of belonging to the statistical sector, the age group of the reference person and the size of the household) of both the initial sample and the realized sample at any stage of the substitution process was compared. Before applying substitution a 64,7% fit between the size and composition of the realized sample and the initial sample could be realized (equals the response rate of the survey without substitution). Especially households with a relatively young reference person, small households and households living in an urban environment turned out to be hard to reach. The substitution process specifically targeted these kinds of households. By replacing non-participating (predominantly young, small and urban) households with similar households it was aimed to reconstruct the composition of the initial sample.

Applying successive (matched) substitutions for non participating households improved the fit to 80,1% (after activating the first substitutes), 85% (second substitutes) and 86.3% (third substitutes). The reverse of this approach is that substitution induces declining participation rates: from 64.7 % of

the initially selected households to respectively 56.9 %, 54.1 % and 50% (third substitutes), indicating a higher reluctance to participate in substitutes.

This experience shows that field substitution must be considered as valuable alternative when conducting surveys. It allows an optimal control during the fieldwork, optimizes the preservation of the designed structure of the sample and is especially fruitful when targeting groups found to have low response rates. A major disadvantage is that it induces higher levels of refusal which results in a supplementary bias in the estimates.

Research Among Ethnic Minorities in the Netherlands

Joost Kappelhof; Social and Cultural Planning office (SCP), Netherlands

In recent years the Social and Cultural Planning Office (SCP) of the Netherlands sponsored two large surveys among ethnic minorities (LAS in 2004 and SIM in 2006). The main goal of the LAS survey was to measure the living conditions of the citizens of ethnic origin in the 50 largest cities of the Netherlands. The SIM Survey was aimed at measuring the socio-economic integration, the attitude of ethnic minorities towards Dutch society and their opinions on the socio-cultural integration of ethnic minorities. The minority groups being investigated in both surveys are the 4 major ethnic groups in the Netherlands (people of Moroccan, Turkish, Antillean and Surinamese descent) and a Dutch control group. In the Netherlands these minority groups are seen as difficult to survey because of cultural differences, language problems and low response rates.

The paper will cover trade-offs we made in the design and the data collection phase of both surveys to ensure the data quality. Furthermore, it will give some useful suggestions about possible strategies to increase response rates and which techniques to avoid. Also we will briefly discuss which areas to give extra attention to when conducting this type of social research.

Issues in the design phase are the (definition of the) target population, the choice of a sampling frame, choice of measurement instrument and mode of measurement. Furthermore we will discuss trade-offs with respect to interviewers, languages of the measurement instrument, type of sample, sample size and expected response rate. The response issue was of particular interest since immediately before start of the fieldwork of the LAS2004 survey we were faced with the tragic assassination of Dutch filmmaker Van Gogh by a young Moroccan radical, followed by a tense atmosphere and a hardened attitude towards Muslims which was expected to have a large impact on the willingness to participate.

Keywords: Ethnicity issues, Methodology, Data quality and Trade-offs

Surveying Victimization Among Immigrant and Ethnic Minority Groups

Jo Goodey; European Union Agency for Fundamental Rights, Austria

Sami Nevala; European Union Agency for Fundamental Rights, Austria

Immigrants and ethnic minorities are some of the most under-researched groups in criminal victimisation survey research. With this in mind, the European Union Agency for Fundamental Rights (FRA) has conducted innovative pilot victimisation survey research in six EU Member States to explore the feasibility of collecting survey data on immigrant and ethnic minority groups. The surveys have consisted of face-to-face interviews on crime victimisation and feelings of safety as well as experiences of and attitudes towards the police. A prerequisite for participation is that the respondents identify themselves as belonging to an immigrant or ethnic minority group.

The pilots have tested two methods for locating the respondents: a random route procedure and a combination of CATI screening and focussed enumeration, where the respondents are asked whether they know of people living in their immediate vicinity who would be eligible to participate in the study. In both cases the sampling was limited to those regions where a high concentration of immigrants or ethnic minorities could be identified based on available population information.

The presentation will consider the merits and limitations of the two fieldwork approaches in the six EU Member States. Given the challenges of ensuring high quality deliverables for cross-national survey

research on 'difficult to survey' groups, the presentation will explore the potential for application of the fieldwork approaches in other European countries with the aim of producing results on these under-research groups.

Labour Market and Social Security: a New Panel Study for Research on German Social Code II

Rainer Schnell; University of Konstanz, Germany

Juliane Achatz; Institute for Employment Research, Germany

Bernhard Christoph; Institute for Employment Research, Germany

Mark Trappmann; University of Konstanz, Germany

Claudia Wenzig; Institute for Employment Research, Germany

The German social security system has been radically transformed during the last years. Since very little is known of the resulting changes in living conditions a new household panel study 'Labour Market and Social Security' (IAB-PASS) was initiated and started in December 2006.

In this presentation we introduce design and implementation of this survey. A scientific use file of wave 1 of this dataset is expected in spring 2008.

The study investigates the life situation of benefit recipients of the recently introduced Unemployment Benefit 2. Among the main analytic interests are:

- A comparison of recipients' economic situation and social circumstances to those of the rest of the German population.
- Providing a database that allows scientists to monitor processes of entering as well as leaving recipient status.
- When looking at processes of leaving recipient status a particular focus is on an evaluation of policy measures aiming to assist recipients to do so.

The targets of the study imply a couple of methodological problems to be solved by the design of the study. Among the main problems to be addressed are:

- Since recipients of Unemployment Benefit 2 are a minority of the German population (below 10 %) and, moreover, some of the policy measures described above will only be administered to a minority of recipients, a massive oversampling of benefit recipients is necessary in order to address the evaluation-goals of the study.
- Benefit recipients are a difficult survey population: High refusal rates, low reachability by landline phone and language as well as comprehension problems have to be anticipated.
- The goal of monitoring processes of entering recipient status asks for inclusion of a sufficient number of candidates for entries (i.e. high-risk-population): As truncated samples are difficult to create, hardly refreshable and coefficients potentially biased, a disproportionately stratified sample with oversampling of high risk households seems more advisable.
- Financial restrictions do not allow for the study to be administered by personal interviewing only.

Solutions to these problems are complex and their discussion will make up for the major part of the presentation. The most important features are:

- The use of two different subsamples, one for the general population and one for benefit recipients. The latter is drawn from administrative data of the German Employment agency.
- The use of a commercial database of building addresses as a sampling frame for the sample of the general population. The address-level status information included in this data set allows a stratification of the sample.

- A mixed-mode approach combining CATI and CAPI was used. It allows for cheap CATI interviews, where possible, while transferring those that cannot be reached by phone to CAPI, thus avoiding a possible selectivity bias caused by under-average accessibility of benefit recipients by landline phone.

Where Have All the Young People Gone?! an Analysis of the Difficulties in Locating and Gaining Access to the Juvenile Segment in Telephone Surveys.

Sara Pasadas Del Amo; Institute for Social Studies of Andalusia, Spain

Micaela Soria Zambrano; Institute for Social Studies of Andalusia, Spain

Manuel Trujillo; Institute for Social Studies of Andalusia, Spain

Marga Zarco Uribe-Echevarría; Institute for Social Studies of Andalusia, Spain

The problem of the no-answer rate in both personal and telephone surveys has been an issue of great concern in the sector for decades. Several studies have shown that the two principal components of this variable, namely the inaccessibility of potential respondents and their refusal to participate in surveys, have evolved in a negative manner. Changing lifestyles, the development of new technologies and saturation caused by the growing demands of the marketing sector and commercial and social research are all factors that serve, in part, to explain the increasing difficulties involved in producing complete samples. However, these factors do not affect the entire population in the same manner. Indeed, the problem of the no-answer rate affects respondents with different profiles in diverse ways. Traditionally, young people have been a difficult population to locate, albeit when contact is made, they are generally more willing to take part in surveys. As regards the issue of location, there are numerous conditioning factors that affect this group to a larger degree such as job instability or geographic mobility. Furthermore, as the literature has pointed out, young people are more difficult to contact once they have been localised due to the fact that they tend to spend less time at home than other types of respondents.

The aim of the present paper is to explore in depth the specific difficulties involved in locating and gaining access to younger populations in survey research. To do so, data is analysed on incidences that have occurred when doing field work for different computer-assisted telephone interviews conducted by the Institute for Advanced Social Studies of Andalusia, as well as other data arising from secondary statistical sources.

Index of people

- Aalbers, T.G. , 123
Abreu Lopes, Cláudia , 30
Achatz, Juliane , 261
Al-Hamad, Alaa , 199
Aleksynska, Mariya , 98
Alexandrowicz, Rainer , 66
Allum, Nick , 88, 89
Aslam, Aqib , 37
- Börsch-Supan, Axel , 22
Bühlmann, Marc , 79, 80
Bachteler, Tobias , 64, 251
Balinskaitė, Violeta , 218
Bandilla, Wolfgang , 153
Bartolomé Peral, Edurne , 85, 97, 98
Batista-Foguet, Joan Manuel , 190–192
Bauer, Gerrit , 19
Beatty, Paul C. , 54
Bechtel, Timothy G. , 134
Beckers, Tilo , 96
Behr, Dorothee , 48
Bell, Alice , 227
Belli, Robert F. , 256
Benítez Baena, Isabel , 60
Bennett-Harper, Sarah , 145
Benocci, Sara , 166
Bensing, Jozien , 107
Bergsma, Wicher , 117
Berkowitz, Susan G. , 71
Berni, Roberto , 166
Bethlehem, Jelke , 147
Beuckelaer, Alain De, 186
Beullens, Koen , 141
Bie , Steven de , 147
Biemer, Paul , 181
Biffignandi, Silvia , 159
Billiet, Jaak , 51, 77, 107, 140, 175–177
Blohm, Michael , 153, 164
Blom, Annelies , 21, 170
Boelhouwer, Jeroen , 39
Boker, Steven , 105
Bormann, Cornelia , 233
- Bosch, Volker , 179
Bouic, Jerelyn , 45
Boyatzis, Richard , 191, 192
Braeken, Johan , 118
Braun, Michael , 100, 101
Brites, Rui , 84, 85, 204, 205
Brown, Gary , 198, 199
Brunton-Smith , Ian , 248
Bryson, Caroline , 227
Buiatti, Eva , 166
- Callegaro, Mario , 178, 255
Campbell, Stephen , 145
Capó, Aina Maria , 209
Cardoso, Inês , 85, 204, 205
Cartwright, Fernando , 39, 137
Casas-Cordero, Carolina , 171
Castiglioni, Laura , 143
Chambers, Ray , 113
Chen, Tao , 58
Christoph, Bernhard , 261
Chrysanthopoulou, Stavroula , 58
Civardi, Marisa , 224
Cobben, Fannie , 125
Coelho, Bernardo , 85
Coenders, Germà , 209
Cole, Jason , 108, 110, 223
Collini, Francesca , 166
Collins, Debbie , 53
Cools, Mario , 240, 243
Coromina Soler, Lluís , 35, 132, 209
Corrado, Luisa , 37
Cosenza, Carol , 54
Couper, Mick , 152, 154
Coutts, Elisabeth , 73
Crone, Matty H. , 68
Croon, Marcel , 116, 117
Cuesta Azofra , María , 74
- Daalmans, Jacco , 150
Dang, Jeff , 223
Davidov, Eldad , 34, 51, 77, 83, 106
Davis, Bryan , 145

- Dawes, Christopher T. , 238
 Dayton, James J. , 180
 Deding, Mette , 258
 Delsing, Marc , 105
 Demarest, Stefaan , 259
 Dept, Steve , 47
 Devine, Paula , 250, 251
 Dijkstra, Wil , 43, 59, 62
 Dijkum, Cor van, 106
 Dmitrović, Tanja , 222
 Dornitz, Jeffrey , 154
 Doorn, Leo van, 144
 Draisma, Stasja , 60
 Driessens, Sabine , 259
 Dubois, Dominique , 110
 Dundler , Agnes , 251
- Eckman, Stephanie , 140
 Edwards, Brad , 12, 16
 Engel, Dirk , 251
 Epland, Jon , 216
 Erens, Bob , 228
 Erkip, Feyzan , 241
 Eva, Gillian , 36, 37, 185
 Eyerman, Joseph , 181
- Fabrizi, Enrico , 159
 Farrall, Stephen , 247, 249
 Fedáková, Denisa , 205
 Ferligoj, Anuška , 207, 209, 210
 Ferreira de Almeida, João , 84
 Feskens, Remco , 257, 258
 Fischer, Jakub , 219
 Fitzgerald, Rory , 14
 Folmer, Henk , 105
 Forni, Silvia , 166
 Forsyth, Barbara , 45
 Fowler, Floyd J. , 54
 Franzini, Luisa , 91
 Frick, Joachim R. , 140
 Fridberg, Torben , 258
 Fromseier Mortensen, Maja , 102
 Fuchs, Marek , 61, 178
 Funke, Frederik , 161
- Göritz , Anja , 160
 Gómez Benito, Juana , 65, 67
 Gabler, Siegfried , 25, 26, 179
 Gajria, Kavita , 110
 Galin, Mickey , 179
 Galindo Garre, Francisca , 68
 Gallagher, Patricia , 103
 Gallhofer, Irmtraud , 43
- Gamrot, Wojciech , 145
 Ganninger, Matthias , 26
 Ganzeboom, Harry B.G. , 196
 Gauckler, Britta , 10
 Gelissen, John , 231
 Gini, Rosa , 166
 Giuliani, Gianluca , 109
 Glasner, Tina , 52
 Glorieux, Ignace , 244
 Gluza, Joanna , 101
 Goicoa, Tomas , 112, 113
 Goodey, Jo , 260
 Grabka, Markus M. , 140
 Graf , Eric , 213
 Graham, Jenny , 227
 Gramlich, Tobias , 147
 Granda, Peter , 17, 18
 Gray, Emily , 247, 249
 Greene, Giles , 184
 Groot, Merel S. de , 15
 Groves, Robert M. , 171, 173
 Guia, Jaume , 210
 Guillén, Laura , 34, 191, 192
- Häder, Sabine , 25, 179
 Hölscher, Michael , 94
 Haag, Antoine , 218
 Hagenaaars, Jacques , 93, 95, 116, 117
 Haller, Therese , 129
 Halman, Loek , 231
 Hank, Karsten , 22
 Hansen, Sue Ellen , 12, 15, 20
 Harkness, Janet , 15, 44, 48
 Harrisson, Eric , 92
 Harsløf, Ivan , 196
 Heeney, Catherine , 227
 Heller, Günther , 234
 Heselmans, Frederic , 86
 Heyden, Johan Van der, 259
 Hidalgo , María-Dolores , 65, 67
 Hill, Craig A. , 181
 Hlebec, Valentina , 208, 209
 Hoffmeyer-Zlotnik, Jürgen H.P. , 68, 78, 195,
 209, 217
 Hohl, Katrin , 248
 Hood, Kerenza , 184
 Hoogendoorn, Adriaan W. , 150
 Hooghe, Marc , 89
 Hopt, Oliver , 252
 Howard, Marc Morjé , 92
 Hughes, Graham , 251
 Hulliger, Beat , 199
 Hyman, Laura , 38

- Hyttiäinen, Pirjo , 20
 Iezzi, D.F. , 136
 Iglič, Hajdeja , 91, 208
 Ingledew, David , 184
 Inglic, Rihard Tomaz , 213
 Isander, Ulf , 180
 Ismail, Zenobia , 135
 Ivans, Rhys , 184

 Jäckle, Annette , 169, 170, 185
 Jürges, Hendrik , 22
 Jackson, Jonathan P. , 247–249
 Jagodzinski, Wolfgang , 93, 95
 Jakobsen, Vibeke , 258
 Jaktlund, Asa , 180
 Jakubowski, Maciej , 102
 Janssen, Björn , 186
 Jaspers, Eva , 237
 Jelonek, Magdalena , 31, 122
 Johnson , Mark , 22
 Johnson, Timothy , 100, 101
 Jowell, Roger , 36, 37
 Juncosa, Sandra , 67
 Jurczak, Anna , 122

 Kaczmirek, Lars , 153
 Kalgraff Skjåk, Knut , 165
 Kaminska, Olena , 176
 Kampen, Jarl K. , 88
 Kankarash, Milosh , 94
 Kappelhof, Joost , 260
 Kawabata, Akira , 231
 Kentoš, Michal , 205
 Kimmie, Zaid , 135
 Kleiner, Brian , 45
 Knežević Cvelbar, Ljubica , 222
 Knoppen, Desiree , 84
 Koch, Achim , 21, 164
 Koelet, Suzana , 244
 Kogovšek, Tina , 208, 209
 Kohler, Ulrich , 171
 Kolsrud, Kirstine , 162, 163
 Kragelj, Boris , 240, 241
 Kraus, Katrin , 51, 128
 Krause, Jürgen , 252
 Krebs, Dagmar , 68, 78, 209
 Kreuter, Frauke , 169, 171, 172
 Krieger, Ulrich , 242
 Kriesi, Hanspeter , 79, 80
 Kristenson, Margareta , 53
 Kritski, Eugene , 137
 Krupnik, Seweryn , 122

 Kuha, Jouni , 248
 Kuhn, Ursina , 237
 Kurth, BM , 234
 Kustura, Marinela , 27

 Laaksonen, Seppo , 25, 26
 Laborde, Caroline , 121
 Laflamme, François , 172
 Laka Mugarza, Jon Paul , 97
 Lalor, Thérèse , 23
 Lamb, Julie , 251
 Lamei, Nadja , 216
 Landmann, Juliane , 202
 Lange, C. , 234
 Lauth, Hans-Joachim , 80
 Leaveron, Amy , 192
 Lee, Jinkook , 18
 Lehmann, Bernard , 129
 Lelièvre, Eva , 121
 Lemay, Michael , 172
 Lensvelt-Mulders, Gerty , 149, 150
 Letki, Natalia , 89
 Levin, Kerry , 45, 145
 Lievens, Filip , 186
 Lipps, Oliver , 60, 170
 Littvay, Levente , 238
 Lobe, Bojana , 70
 Longford, Nick , 126
 Loosveldt, Geert , 140, 150, 176
 Lozar Manfreda, Katja , 158
 Lubbers, Marcel , 237
 Lugtig, Peter , 146, 149, 150
 Luijkx, Ruud , 93, 95
 Lyhagen, Johan , 128
 Lynch, David , 243
 Lynn, Peter , 25, 169, 170, 185

 Müller, Walter , 19
 Macchia, Stefania , 133
 Madans, Jennifer , 10
 Maggino, Filomena , 119, 120
 Magidson, Jay , 223
 Maitland, Aaron , 10
 Makovec Brenčič, Maja , 222
 Mali, Franc , 210
 Manabe, Kazufumi , 231
 Marchand, Miquelle , 155
 Martín, Begoña , 199
 Martín, Montse , 67
 Martínez-Herrera, Enric , 76
 Martini, Maria Cristiana , 219
 Mascherini, Massimiliano , 194
 Masiulaityte, Inga , 218

- Matelič, Uroš , 210
 Matschinger, Herbert , 66
 Matsuo, Hideko , 176
 Maydan, Mike , 172
 Mazaheri, Mehrdad , 128
 Mazouch, Peter , 219
 Mceachern, Steven , 242
 McGee, Alice , 53, 255
 Menold, Natalja , 61
 Mesch, Gustavo S. , 92
 Meuleman, Bart , 28, 29, 77
 Meulemann, Heiner , 230, 232
 Milfelner, Borut , 223
 Militino, Ana Fernández , 113
 Miller, Kirsten , 8–10
 Mitina, Olga , 51
 Mohler, Peter Ph , 167
 Moons, Elke , 243
 Moor, Nienke , 232
 Moors, Guy , 94
 Morren, Meike , 73
 Mугan, Güiliz , 241
 Mumel, Damijan , 223
 Munda, Giuseppe , 38
 Murgia, Manuela , 50, 133
 Museaux , Jean-Marc , 212
 Mysíková, Martina , 213
- Németh, Renáta , 117
 Nardo, Michela , 38, 194
 Nascimento Silva, Pedro Luis do , 199
 Need, Ariana , 232
 Neubarth, Wolfgang , 153
 Nevala, Sami , 260
 Nicolaas, Gerry , 173, 226
 Nicolet, Sarah , 239
 Nicols, Simon , 235
 Niemi, Iiris , 20
 Nofuentes, Santiago , 67
 Noorlind-Brage, Helle , 53
 Norberg, Alicia , 45
- O'Brien, Jennifer , 145
 O'Muircheartaigh, Colm , 140
 Oberski, Daniel , 35, 42, 43, 160
 Ograjenšek, Irena , 221, 222
 Olson, Kristen , 172
 Ongena, Yfke P. , 43, 59, 62
 Oort, Frank van , 105
 Ormston, Rachel , 227
 Orten, Hilde , 167
 Oud, Han , 104–107
- Pääkkönen, Hannu , 20
 Padilla Garcia, Jose Luis , 9, 60, 65, 67
 Pan, Yuling , 45
 Paniagua San Martín, Milagros , 216
 Pasadas Del Amo, Sara , 74, 262
 Pasquini, Jacopo , 166
 Patulny, Roger , 89
 Pavlova, Veselka , 57
 Pennell, Beth Ellen , 21, 168
 Pereira, Cícero , 77
 Persson, Andreas , 52
 Petrakos, George , 56, 58
 Pettersson, Thorleif , 231
 Peytchev, Andy , 124, 125, 172, 181
 Peytcheva, Emilia , 13, 125, 173
 Piani-Acosta, Giorgina , 46
 Pichler, Florian , 97
 Pintaldi, Federrica , 109
 Pisnik Korda, Aleksandra , 223
 Pollak, Reinhard , 19
 Popa, Clara L. , 90
 Przybysz, Dariusz , 142
 Puente, Manuel de la , 15
- Quaglia, Martine , 23, 72
 Quandt, Markus , 165
- Rószkiewicz, Małgorzata , 121
 Raczka-Vargas, Danuta , 213
 Raeside, Robert , 58
 Rangunathan, Trivellore , 172
 Rammstedt, Beatrice , 120
 Ramos, Alice , 77, 84
 Reeskens, Tim , 29, 89
 Reiher, Jörg , 64, 251
 Reips, Ulf-Dietrich , 161
 Riis, Ole , 232
 Roberts, Caroline , 183–185
 Robins, Cynthia S. , 71
 Robling, Mike , 184
 Roe, David , 181
 Rudas, Tamás , 117
 Ruiten, Stijn , 231
 Russell, Ian , 184
 Rybkowska, Anna , 213
- Sánchez Carrión, Juan Javier , 148
 Saei, Ayoub , 114
 Saisana, Michaela , 39
 Sajitha , OG , 24
 Sala, Emanuela , 254, 255
 Saltelli, Andrea , 38
 Sander, Lesley , 184

- Saris, Willem E. , 32, 33, 35, 43, 81, 160, 191
 Satorra, Albert , 127
 Sauer, Juliane , 179
 Sawiński, Zbyszek , 165
 Sayers, Adrian , 184
 Scheidt-Nave, , 234
 Scherpenzeel, Anette , 153
 Schmeets, Hans , 257, 258
 Schmidt, Peter , 77, 106
 Schneiderat, Götz , 179
 Schnell, Rainer , 63, 64, 73, 143, 171, 251, 252, 261
 Scholz, Evi , 194
 Schouten, Barry , 186
 Schröder, Mathis , 22
 Schrooten, Hilde , 245
 Sciarini, Pascal , 239
 Segovia Guisado, José Manuel , 148
 Serlavos, Ricard , 191, 192
 Seykova, Svetlana , 74
 Shaw, Chris , 184
 Shipp, Stephanie , 145
 Sienkiewicz, Robert , 145
 Sierminska, Eva M. , 140
 Silvestre Cabrera, María , 97
 Smetanová, Tamara , 213
 Smit, Johannes H. , 60
 Smith, Patten , 88
 Smith, Tom W. , 139, 141
 Snoj, Boris , 223
 Sonck, Nathalie , 150
 Sonneveld, Martijn , 129
 Soria Zambrano, Micaela , 262
 Spinney, Jamie , 244
 Statham, Paul , 203
 Stathopoulou, Theoni , 205
 Stempfhuber, Max , 252
 Stocké, Volker , 133, 164, 208
 Stoel, Reinoud D. , 68
 Stoop, Ineke , 140, 147, 163, 175, 201, 202
 Stoykova-Kanalieva, Andreana , 252
 Stringfellow, Vickie L. , 103
 Sturgis, Patrick , 87–89
 Sztabinski, Franciszek , 142
 Sztabinski, Pawel , 142
 Szukielojc-Bienkunska, Anna , 213

 Törmälehto, Veli-Matti , 216
 Tafforeau, Jean , 259
 Talmud, Ilan , 92
 Tam, Tony , 90, 101
 Tampubolon, Gindo , 109
 Taylor, Rebecca , 226

 Tejero Cabello, Palmira , 67
 Theuns, Peter , 128
 Thompson, Frances , 45
 Tijdens, Kea , 159, 218
 Todosijevic, Bojan , 111
 Toharudin, Toni , 107
 Torres, Analia , 85, 204, 205
 Tourek, Stepán , 213
 Trappmann, Mark , 143, 261
 Trechsel, Alexander , 81
 Trujillo, Manuel , 262
 Tuerlinckx, Francis , 118
 Tutkun, Aysel , 129, 238
 Tzavidis, Nikos , 113

 Ugarte Martínez, M. D. , 113
 Uhrig, Noah , 254, 255
 Ultee, Wout , 232
 Uskül, Ayşe , 13

 Väisänen, Paavo , 20
 Výrost, Jozef , 205
 Vaart, Wander van der, 52
 Vala, Jorge , 77, 84
 Vandecasteele, Leen , 177
 Vangeel, Koen , 245
 Vehovar, Vasja , 142, 158
 Veld, William van der, 33, 81, 236
 Verhagen, Josine , 203
 Verheul, William , 107
 Verhue, D. , 123
 Verma, Vijay , 215
 Vermeylen, Greet , 72
 Vidoni , Daniele , 194
 Villar, Ana , 15, 44, 48
 Vis, Corrie , 154
 Visser, Hans , 123
 Vivier, Géraldine , 121
 Voller, Fabio , 166
 Vries, Jannes de, 196
 Vringer, Kess van, 123

 Wagner, James , 172
 Wagner, Simone , 135
 Wardrop, Tim , 23
 Warner, Uwe , 195, 217
 Weiss, Felix , 19
 Welzel, Chris , 81
 Wenemark , Marika , 53
 Wenzig, Claudia , 261
 Wernli, Boris , 237
 Wets, Geert , 243
 Weziak, Dorota , 121

White, Clarissa , 227
Widdop, Sally , 185
Williams, John , 184
Willis, Gordon , 8, 9, 46
Wirth, Heike , 19
Wolf, Christof , 17, 193

Yan, Ting , 172
Yang, Keming , 81

Züll, Cornelia , 202
Zabkar, Vesna , 222
Zadło, Tomasz , 114
Zandberg, Izabella , 71
Zarco Uribe-Echevarría, Marga , 262
Zavarrone, Emma , 224
Zelený, Martin , 213, 215
Ziese, Thomas , 234
Ziherl, Petra , 208
Ziniel, Sonja , 73
Zouwen, Johannes van der, 60
Zuev, Dennis , 196
Zupaniè, Tina , 142
ZuWallack, Randal S. , 180, 181
Zwingenberger, Anja , 252