ECONSTOR
Make Your Publications Visible.

A Service of
ZBW
Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Blom, Annelies G.; Lynn, Peter; Jäckle, Annette

Working Paper

Understanding cross-national differences in unit non-response: the role of contact data

ISER Working Paper Series, No. 2008-01

Provided in Cooperation with:
Institute for Social and Economic Research (ISER), University of Essex

Suggested Citation: Blom, Annelies G.; Lynn, Peter; Jäckle, Annette (2008) : Understanding cross-national differences in unit non-response: the role of contact data, ISER Working Paper Series, No. 2008-01, University of Essex, Institute for Social and Economic Research (ISER), Colchester

This Version is available at:
https://hdl.handle.net/10419/92001

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

ISER Working Paper Series

Understanding Cross-National Differences in Unit Non-Response: The Role of Contact Data

Annelies G. Blom
GESIS-ZUMA, Mannheim

Peter Lynn
Institute for Social and Economic Research
University of Essex

Annette Jäckle
Institute for Social and Economic Research
University of Essex

No. 2008-01
February 2008

INSTITUTE FOR SOCIAL
& ECONOMIC RESEARCH

E·S·R·C
ECONOMIC
& SOCIAL
RESEARCH
COUNCIL

www.iser.essex.ac.uk

Non-technical summary

We briefly review the causes and consequences of non-response and aspects specific to cross-national studies. Non-response can occur at different stages: because sample members cannot be located, because they cannot be contacted, or because they refuse or are unable to take part in the survey. If the responding sample members are not a random sub-set of the initial sample, then estimates may be biased. If, however, the differences between respondent and non-respondent sample members can be fully explained by auxiliary information known for both groups, then unbiased estimates can be produced by using statistical adjustments. In cross-national research, the issue is further complicated, since both the factors explaining survey participation and the auxiliary information available may be different across countries.

Contact data are a source of information about both respondents and non-respondents, which measures some of the aspects of the respondent recruitment process. Contact data refer to information recorded by interviewers about their attempts at contacting sample members. They typically include information about each contact attempt made, such as the day, time, mode and outcome of each call. Contact data are primarily generated and used for survey management purposes – initially for interviewers to organise their workload, but also for the survey organisation to monitor fieldwork procedures and interviewer efforts. In recent years it has been realised that these process data are also useful for methodological analyses. As a result, increasing numbers of survey organisations are collecting call record data in a systematic way and in some cases releasing these auxiliary data into the public domain.

We propose a typology of the processes generating contact data and discuss the implications for their utility for methodological research. We further propose and discuss four criteria for the utility of contact data for cross-national analyses: 1) equivalence of design, 2) equivalence of implementation, 3) equivalence of coding fieldwork outcomes and 4) data availability. We then review existing studies of non-response in cross-national surveys. Early studies used information at the survey level to compare response rates and trends in different countries. More recent studies use individual-level call record data to examine the processes leading to survey co-operation. This is however still an emerging field and the small number of cross-national surveys for which contact data exist and the small number of empirical studies that have been carried out to date mean that empirical facts about non-response in cross-national surveys have not yet emerged. We end with a discussion of the potential ways in which contact data can be used for non-response adjustment and fieldwork specifications and with recommendations for the collection of contact data on cross-national surveys.

Understanding Cross-National Differences in Unit Non-Response: The Role of Contact Data

Annelies G. Blom, Peter Lynn and Annette Jäckle

February 2008

ABSTRACT

This paper showcases ways in which contact data can provide useful auxiliary information to study non-response and devise new strategies for statistical adjustment and fieldwork specification in cross-national surveys. We propose a typology of the processes generating contact data and discuss their implications for the utility for methodological research. We also propose criteria for the comparability of contact data in cross-national surveys. To illustrate the potential of contact data, we review existing studies of non-response in cross-national studies. The paper ends with recommendations for cross-national studies, with the aim of increasing the widespread availability and use of standardised contact data.

Keywords: call records, para-data, cross-national surveys, comparative research

JEL codes: C42

Acknowledgement:
This paper will form the basis of an invited presentation at the 3MC conference in June 2008 (http://www.3mc2008.de) and a revised version will appear as a chapter in the monograph book of the conference, to be published by John Wiley. We would like to thank Lars Lyberg for helpful suggestions. Part of this paper is based on work carried out during Annelies Blom's visit to the European Centre for Analysis in the Social Sciences (ECASS) at the Institute for Social and Economic Research, University of Essex, supported by the Access to Research Infrastructure action under the EU Improving Human Potential Programme.

Contact:
Annelies G. Blom, GESIS-ZUMA, Postfach 12 21 55, 68072 Mannheim, Germany. Email: annelies.blom@gesis.org; Peter Lynn, Institute for Social and Economic Research, University of Essex, Colchester CO4 3SQ, UK. Email: plynn@essex.ac.uk; Annette Jäckle, Institute for Social and Economic Research, University of Essex, Colchester CO4 3SQ, UK. Email: aejack@essex.ac.uk.

1. Introduction

The analysis of cross-national survey data can be hindered by non-response. It is not uncommon for cross-national surveys to have very different response rates in different countries. This raises awareness amongst analysts of the potential for differential non-response errors, which might bias estimates of differences between countries. Analysts are sometimes left wondering whether the countries with the lowest response rates should be excluded from their analysis altogether.

In this paper, we review the role contact data can play on cross-national surveys in helping to understand the nature of non-response, to identify appropriate field procedures regarding non-response and to adjust for non-response. The focus of the paper is therefore defined by two dimensions. The first is the context of non-response research on cross-national surveys. The second is the focus on the use of contact data for non-response research. The intersection of these two dimensions – the use of contact data in studies of non-response on cross-national surveys – is an emerging field in which little research has been carried out to date. We review existing studies but we find that there remain considerable challenges in collecting and analysing contact data on cross-national surveys.

To place the discussion of contact data in context, we begin (section 2) with a brief review of survey non-response, its components and its impacts. We highlight those aspects that are particular to the case of cross-national surveys. We also describe the role of auxiliary data in general and consider contact data as an additional category of auxiliary data, albeit one with special characteristics.

In section 3 we introduce a typology of contact data and discuss the relationship between the process generating the data and the utility of the data to researchers. In section 4 we suggest four criteria for the usefulness of contact data in the particular context of cross-national surveys. We then summarise existing research on cross-national comparisons of non-response, highlighting the distinction between studies carried out without the benefit of contact data and those that have used contact data (section 5).

The paper concludes (section 6) with a discussion of the potential role of contact data in new strategies for non-response adjustments and in field work specification and control on cross-national surveys. Recent progress in this area is reviewed and a number of recommendations are made for the collection of contact data on cross-national surveys with the goal of making progress towards the widespread availability and use of standardised contact data. It is our belief that contact data can make a very useful contribution to the

treatment of non-response on cross-national surveys, but that to realise this potential both the conceptualisation of the data to be collected and the procedures to collect them need to be standardised in important ways.

2. Non-Response on Cross National Surveys

Survey Non-Response

Survey non-response is the absence of (some) data for (some) survey sample units. In other words, a survey may be designed to collect a certain set of data items from a certain well-defined sample of units but, for a host of reasons, it may not always prove possible to obtain all of the data items for all sample units (Groves et al 2004, chapter 6). There may be some sample units for which no data at all can be collected. This is referred to as *unit non-response*. There may be other units for which some data, but not all data items, are collected. This is referred to as *item non-response*. In this paper we focus solely on unit non-response which from hereon we shall refer to simply as 'non-response'.

The process that leads to non-response is typically a complex one. The specific components of the process may vary between surveys, but usually there are three main stages in the process. These are worth distinguishing as they may have rather different impacts on bias (and the extent to which this is true may differ between estimates) and consequently different implications for how bias may best be treated. The three stages are *location*, *contact* and *co-operation*. The location stage involves the survey organisation locating each sample unit. For example, for a telephone survey this involves obtaining a correct telephone number for each sample member; for a face-to-face interview survey this involves both obtaining a correct address and successfully locating it in the field. The contact stage involves making direct contact with the sample unit in the same medium as the survey interview. For a telephone survey this involves talking on the telephone to the sample member; for a face-to-face survey this involves talking face-to-face with the sample member. The co-operation stage involves successfully obtaining the desired survey data, usually either through an interview or a self-completion questionnaire.

Response is only obtained if the location, contact and co-operation stages are all passed successfully. It should be obvious that the processes leading to success are different at each stage. Consequently, the correlates of success are likely to be different at each stage. These correlates will include both characteristics of the sample units (e.g. Lynn et al 2002) and characteristics of the field work strategies adopted by the survey. For example, the

propensity to make contact will (in the case of a face-to-face survey) depend on characteristics of sample members such as the amount and distribution of time spent at home, which will in turn be influenced by employment patterns, leisure activities, and so on, but also on characteristics of the field work strategy such as the number and timing of call attempts (Groves and Couper 1998). Co-operation may be influenced by a different set of factors again, such as social integration, attitudes towards authority and disposition towards the interviewer on the one hand and interviewer introductions, survey materials, incentives and so on the other (Groves et al 2000).

The main reason that researchers are interested in non-response is that it may bias estimates based upon the survey data, perhaps leading to misleading conclusions. An unadjusted estimate will be unbiased only if non-respondents do not differ systematically from respondents in terms of the survey items that contribute to the estimate. This situation is referred to by Little and Rubin (2002) and others as *missing completely at random* (MCAR). If non-respondents *do* differ systematically from respondents then estimates will be biased (*not missing at random* - NMAR) unless these differences can be fully explained (in a statistical sense) by other available survey items, in which case unbiased adjusted estimates can be produced (*missing at random* – MAR).

Let us dwell for a moment on this definition of non-response bias as it is central to the discussion of this paper. First, note that bias is specific to each estimate. Thus, an identical non-response process can lead to bias in one estimate but not in another. This can result in different estimates from the same survey (with, by definition, the same unit non-response process) having very different bias properties. Second, note that bias depends on the other data items available for sample members. This could result in two surveys with identical non-response characteristics producing differently-biased estimates of the same parameter. This would happen if one of the surveys includes additional items that fully explain differences in non-response propensity between sample members, while the other does not. These additional items could either be other survey items (available only for respondents), or extra data that are available for all sample members (respondents and non-respondents). The latter are a subclass of auxiliary data (which can also include aggregate data). Auxiliary data that directly measure aspects of the survey data collection process have become known as para-data (Couper 2005). Couper and Lyberg (2005) distinguish between macro- and micro-level para-data. The former are survey process indicators measured at the whole-sample level (e.g. response rate or length of fieldwork period), while the latter are those measured for each sample member (e.g. imputation flags or call records). Micro-level auxiliary data that are

available for both respondents and non-respondents – including micro-level para-data – can be more powerful in explaining non-response than survey variables, as the latter are only available for respondents. The key is to find "variables that are related to both the probability of response as well as to key survey outcomes" (Kreuter et al 2007). Potential sources include:

- The sampling frame (e.g. sex and age if the frame is a population register);
- Linked micro-level data (e.g. publicly-available administrative data that can be linked using identifiers that are on the frame, such as full names);
- Linked geographical or other aggregate-level data (e.g. Census small area data or other data that can be linked via grid reference or postal code);
- Interviewer observations (e.g. regarding the sampled dwelling or the neighbourhood);
- Interviewer characteristics (e.g. sex, age, years of experience);
- Contact data;
- Other survey process data.

The challenge for the researcher is to identify a set of auxiliary data that will transform a NMAR process to a MAR one. In other words, non-response bias can be removed from estimates if additional data items are available that correlate both with the survey estimate and with the propensity to respond. The bias can only be removed entirely if the additional items explain all of the difference between respondents and non-respondents in the estimate. But to approach this situation, the additional items will need to explain all three stages of the non-response process. This is why para-data, and particularly contact data, may be a useful type of auxiliary data. Contact data may supplement other forms of auxiliary data as they may be particularly well-suited to explaining the location and contact stages of the non-response process.

Cross-National Surveys

Cross-national surveys differ from national or sub-national surveys in a number of important ways (Lynn et al 2006). There are likely to be design differences between countries that impact on the non-response process: for example, the availability of different kinds of sampling frames in different countries (Lynn et al 2006; Lipps and Benson 2005). With respect to implementation, a key consideration is that data collection is typically organised at

the national level, with a different survey organisation involved in each country. This potentially introduces *house effects* (Smith 1978 1982) that are confounded with country effects. There may be differences between survey organisations in the types of persons recruited to work as interviewers, in the interviewer payment structure and so on. This, and cultural differences, may lead to different interviewer practices in different countries, such as differences in working hours/days (Lipps and Benson 2005). In summary, there are likely to be differences between countries in field practices. There are also likely to be differences in the population distribution of characteristics associated with non-response (for example, the proportion of persons living in households where all adults have a full-time job). And the relationship between those characteristics and the propensity to respond may differ between countries. This can happen, for example, because full-time jobs have different implications for at-home patterns in different countries (e.g. average commuting times can be very different) or because interviewers have different typical working hours in different countries. What is unclear is the extent to which differences between countries in field practices will produce differences in non-response bias.

With respect to analysis, a central aim of most cross-national surveys is comparison between countries. Thus, key survey estimates are estimates of differences between domains. To make unbiased comparisons between countries we essentially need one of two conditions to hold. Either the non-response process must be MCAR or we must identify - and use appropriately - a set of additional items that turn the process from NMAR to MAR. A key issue for this paper is the extent to which contact data can contribute towards this set of additional items. However, for the reasons stated above, there is no reason to suppose that the same set of items should work best for every country, nor that particular items will have the same relationship with non-response in every country.

There are of course other sources of error that may vary between countries, such as coverage error, sampling error and measurement error. These too can affect the analysis objectives of cross-national surveys. They may also interact to some extent with non-response error. In other words, measures to reduce one source of error could potentially increase another source of error. The relationship between sources of error could itself differ between countries and to some extent this could justify differences in survey procedures. Such possibilities must be borne in mind when considering standardisation between countries.

3. Contact Data

Contact data, also known as call-record data, are measurements of key aspects of the process that leads to a fieldwork outcome. They provide information on all sample units, i.e. respondents and non-respondents, and on all contact attempts. Information typically collected includes the date, time, mode (phone, mail or in-person), interviewer and outcome (no contact, interview, refusal, unable, ineligible, appointment, etc) of each contact attempt. The data are used to monitor and optimise different stages of the data collection process. It is not our objective here to discuss in detail how contact data are used during field work. However, the way they are used typically determines the form in which they are collected and hence the form in which they are available for non-response analysis. We therefore provide an overview of the main considerations that influence the form and detail of contact data available to the researcher, in order to set the context for our discussion of the implications for non-response research.

At the case level, the recorded status and contact history is used, typically by the interviewer, to decide when and how to make future calls. At the interviewer level, contact data can be used by the survey organisation to monitor interviewer performance. At the survey level, contact data can be used to monitor where fieldwork progress lags behind schedule, which types of sample units are underrepresented in the achieved sample and should receive more effort, or whether refusal conversion strategies are effective.

Figure 1: Example of a contact form

Respondent ID number	

Number of contact attempt	Date (DD/MM)	Time (24 hr clock)	Mode of visit (code)	Outcome of visit (code)	Notes
1	/	:			
2	/	:			
3	/	:			
4	/	:			
5	/	:			
6	/	:			
7	/	:			
8	/	:			
9	/	:			
10	/	:			

Contact data are collected by the interviewer usually by means of some kind of contact form. Figure 1 shows an example of what the relevant part of a contact form may look like. The outcome code schema for contact attempts must be specified by the researcher and will depend on the survey objectives, the population researched and the sampling frame. (See Lynn et al (2001) and AAPOR (2006) for a review of outcome codes in various survey situations.) The schema need not be complex: the European Social Survey uses just eight general outcome codes.

Paper-based versus computerised fieldwork systems

On any survey, interviewers know the characteristics and outcomes of all calls they have made. Surveys however differ in whether and how this information is captured. Face-to-face surveys historically use *paper-based systems*, where interviewers record the information about the contact process using pen and paper. Paper-based systems are often favoured for face-to-face interviews, even if the interview is carried out using CAPI. This may be because the technology offers greater speed and flexibility in the field, where contact data are often recorded on the doorstep or in other public places and where the interviewer may be moving rapidly from one case to another. Some organisations still use paper-based systems for telephone surveys, though this may be more likely when the interviewing is decentralised than when a centralised facility is used. Centralised CATI surveys, where cases are worked by a number of different interviewers, typically use *computerised systems*, which can vary greatly in capability and complexity. Relational data-base systems are most frequently used, where interviewers enter a call outcome code for each contact attempt according to a code frame specified by the researcher. Date and time are recorded automatically by the system. Computerised systems are increasingly also used for face-to-face surveys, though the data entry by the interviewer may not always be instantaneous. This may take place some time after a call attempt is made, or even just once at the end of each field trip. Instantaneous entry of contact data is less important for face-to-face surveys as only one interviewer is assigned to work on each sample case at any one time.

Case, interviewer and survey fieldwork management

How the data are captured depends, in part, on how they are used. The usage also determines whether contact data are recorded as *call-level* data or summarised and recorded as *case-level* data. For case management, face-to-face interviewers may use full call-level data about each past call (which could be either paper based or computerised) to plan their workload and

decide on the next action for each case. In a centralized CATI setting, computerised data about past calls may provide the input for automated calling schedules, which are algorithms that trigger the next call to each case. The software for computerised contact forms "usually includes a report that summarizes the most recent disposition field in the sample data base" (McCarty 2003, p.398/9). In some fieldwork management systems, the information regarding a call attempt is only stored until the next call attempt is made. At that point, the previous information is overwritten by the outcome of the new call. The final outcome of a series of contact attempts is therefore determined by the outcome of the last contact attempt. No full record of the contacting process is stored and available for later analysis. This kind of data is often produced by systems based on relational data bases (McCarty 2003). Alternatively, the case management may not record the outcome of past calls at all, but instead record the current status of a case and what the next action should be. Action code systems, for example, record whether further calls are to be made or whether supervisors need to decide on the next step (AAPOR 2006). With this type of system the history of call characteristics and outcomes is lost and cannot later be retrieved for analysis. For the purposes of survey management, contact data are often analysed after the completion of fieldwork if the data are not available to the survey organisation electronically and in real-time. The lessons learnt are then applied to subsequent surveys. Recently, analyses of contact data and field interventions have been tested in real-time systems, so-called responsive designs. Interviewers record full contact data on computerised systems and at regular intervals transfer these to the survey organisation together with the survey data. Changes to the survey design are made during fieldwork, based on real-time information about fieldwork and survey outcomes which affect costs and errors (Groves and Heeringa 2006).

Call-level versus case-level data

The survey management processes used by a particular organisation determine whether and at which stage the call-level contact data are reduced to a summarized case-level form: 1) interviewers may return only case-level contact summaries to the survey organisation, 2) the survey organisation may summarize full contact data before releasing the data, or 3) researchers may derive their own summary measures from call-level contact data. In the first case, it is left to the interviewer to derive the required indicators. Typically the indicators may include only a final case outcome (e.g. interview, refusal, non-contact, ineligible) and the total number of calls/visits made. Interviewers are only expected to return these summary indicators for each case they were assigned, even though they may well have recorded more

information in the course of carrying out the field work. Consequently, no information on intermediate contact attempts is stored and available for later analysis. Even if survey organisations do receive call-level contact data from the interviewers, they might decide to code call outcomes into final case outcomes. This can happen for a variety of reasons including data protection issues, commercial sensitivity or because the researcher or sponsor prefers to receive summarised contact data. Finally, methods researchers may code call-level response outcomes into case-level outcomes in order to derive a final outcome code for each sample unit and calculate response rates using their own preferred definitions.

Deriving case outcome codes

Irrespective of the stage at which the case outcome is coded, there are three main methods by which the code may be derived: *most-recent*, *priority* and *subjective* coding. With most-recent coding, the outcome of the last call to a sample unit is defined as the case outcome (AAPOR 2000). Priority coding, on the other hand, involves arranging call outcomes to a priority ordering, in which some outcome codes take priority over others. For example, one would define that achieving an interview takes priority over a refusal, which in turn takes priority over a non-contact (for details on priority coding see Lynn et al 2001). A situation in which an interviewer tries to convert an initial refusal, yet never manages to make contact again, would be coded differently in the two coding systems. If the last call outcome defines the final disposition, this would be a non-contact. According to a priority coding system, it would be a refusal. Finally, subjective coding refers to situations where the rules for converting call outcomes to a case outcome are not defined. Typically in such situations, only descriptions (which can vary in their precision) of each case outcome code are provided. It is left to the coder to decide how to allocate cases to outcomes. This kind of coding is perhaps most common when interviewers are asked to return case-level codes to the survey organisation, though it may also be used by survey organisations carrying out in-office coding.

Implications for non-response research

The contact data resulting from the survey management processes can be used *ex post* for methodological research and is therefore the main focus in this paper. Contact data can be used for non-response research, because they measure aspects that are distinct from those measured by other auxiliary data, suggesting that they may well add explanatory power to that which can be obtained from other auxiliary data alone. They provide data on respondents

and non-respondents and thereby they permit analysis both of the association between contact characteristics and response propensity and simultaneously of the association between contact characteristics and survey estimates.

From a researcher's point of view some types of contact data are more useful than others. Firstly, full call-level data are preferable since a) researchers can choose how to define the final outcome code based on the sequence of call outcomes, b) researchers can create a range of other case-level indicators, for example concerning sequences of call attempts or timings of attempts, and c) researchers can carry out analyses at the call level in addition to analyses at the case level. Secondly, case outcome codes defined by the survey organisation are preferable to case outcomes defined by interviewers, since there is less scope for variations between interviewers and the definitions applied to different cases are more likely to be comparable. Thirdly, contact data generated with computerised systems are preferable to those generated with paper-based systems. The data are likely to be of better quality, as routing and edit checks can be built into the script, reducing the potential for interviewer errors. In addition, the data are more likely to be available for analysis, as they already exist in electronic form and do not have to be keyed first, as would paper-based data.

Despite contact data being a valuable source of information for both survey practitioners and survey methodologists, the methodology for defining, collecting and recording contact data is underdeveloped. We still lack best practice and coherence in (1) the design of contact forms to collect the data, including the technology used for their collection, structure and content; (2) the implementation of contact forms, i.e. instructions given to interviewers regarding how they should fill in the forms; and (3) the coding of contact data, specifically how best to code individual outcomes of call attempts into a final outcome. Without a common understanding on these issues comparability of contact data cannot be achieved. This is a major obstacle to their use on cross-national surveys where comparability is a main objective. The following section points to issues that need to be considered regarding the comparability of contact data. Subsequent sections use the term 'contact data' to refer to call-level contact data available to the researcher.

4. Contact Data for Comparative Research

The growing availability of contact data opens up new possibilities for non-response research. However, in order to compare contact data across different surveys, measurement equivalence is required, as it is for other types of survey data. To achieve this, the contact

data should be input harmonised. It is high time that authoritative international standards for the design and implementation of contact forms and coding of contact data were formulated. Leaning on standards for substantive measurement this section proposes four criteria for such standards for contact data. Achieving equivalence on all four of these criteria should enable high-quality comparative non-response analyses based on contact data.

1) Equivalence of design

For comparative analyses of contact data both the content and the structure of the contact forms need to be equivalent across surveys. Regarding the content of contact forms at a minimum the code frame for the intermediate call outcomes should be the same. This also includes the eligibility criteria (and ineligibility codes). For detailed comparative analyses, the more information is collected in the contact forms of each survey, the more comprehensive the analysis can be. Therefore, in addition to equivalence of the call outcome, equivalence in the collection of date and time, mode of call and other additional variables is desirable.

Regarding the structure of contact forms not only the information collected, but also the way it is collected should be the same across surveys. From questionnaire design methodology we know that differences in question formulation, question format and translation can have a serious impact on the data. Though no such evidence yet exists for contact forms, keeping the structure of contact forms the same should allow maximum comparability.

2) Equivalence of implementation

Next to the design, the comparability of the data is likely to be influenced by the implementation of contact forms in the field. Therefore, it might be desirable to standardise contact form implementation. This includes (1) whether contact data are collected on the computer or on paper, (2) how interviewers are trained and briefed on filling in the forms, and (3) what the fieldwork control procedures are for the contact forms.

The second and third points are critical here. Frames of disposition codes are often difficult to apply to real fieldwork situations. On the doorstep it might for example be difficult to decide, whether a person saying that he/she has no time is a refusing to participate or inviting the interviewer to come back another day. Likewise, in cases where there are language problems a rejection might be a refusal or a genuine inability to participate because of the language.

In addition, interviewers often regard contact forms as unimportant, since they do not concern the main questionnaire and they may perceive that they are not being paid for completing them. Without consistent quality checks on the contact data, they are likely to suffer from negligence. Unfortunately, detailed interviewer training on contact forms and quality control during fieldwork is still the exception. In section 5 we illustrate that even in the case of standardised contact forms researchers are frequently faced with problems of data quality. Due to high levels of missing data for example, detailed analysis of contact data can be hindered.

In addition to problems of comparability *across* surveys, the lenient implementation of contact data collection can be a source of error *within* a survey through interviewer effects. Detailed interviewer training and close monitoring of interviewer performance, in combination with an *ex post* analysis of interviewer effects are key to achieving low levels of variation.

3) Equivalence of coding fieldwork outcomes

As described in section 3, researchers with call-level contact data at their disposal need to decide on how to code the intermediate call outcomes of each sample unit into a final case disposition code for the sample unit. This sounds straightforward, however, since it concerns a *series* of outcomes at sequential calls, defining a case outcome can be intricate. We have discussed two common ways of coding final outcomes: most-recent and priority coding.

When comparing different types of coding schemes McCarty (2003) found only a small impact on overall response rates. This however is not surprising, since for calculating overall response one needs common coding of interviews and eligible sample units only. These are less likely to be affected by coding schemes than contact or refusal rates. In a paper examining the effects of different coding schemes on final outcomes and rates Blom (in progress) looks at the impact of coding on cross-national differences.

4) Data availability

In their analyses of the cross-national decline in response rates Couper and de Leeuw (2003) noted that “for valid cross-cultural and cross-national comparisons, it is of utmost importance that the various sources of nonresponse are reported.” Their comment was made, when access to comparable cross-national contact data was highly unlikely. Nowadays, we can take this a step further and state that for valid cross-cultural and cross-national comparisons access to comparable call-level contact data is of utmost importance. These, in combination with other

types of para-data on survey implementation, enable researchers to code response outcomes and conduct response analyses themselves according to strategies that they regard as optimal.

Considering these equivalence criteria high-quality comparative contact data can be achieved. A note of caution is in place though when analysing the data. Even in the ideal case of perfect equivalence of design, implementation and coding of contact data and access to them cross-national differences in available sampling frames across countries lead to necessary differences in the design of contact data and the distribution of response outcomes. This means that countries with an individual sampling frame drawn from a register will necessarily have slightly different call outcome codes than countries with a household or address-based sample design. More importantly, with different sample designs the distribution of the outcomes are bound to be different, as are the fieldwork processes leading to them. For cross-national research, differences in sampling frames and designs must therefore be taken into account.

However, in many surveys and many countries contact data do not live up to these standards. Lacking information from experimental designs, there is unfortunately insufficient research to estimate what the consequences of not achieving equivalence standards are for the comparability of response analyses. Therefore it is important to be aware of possible differences across countries in the quality and comparability of contact data, when analysing and comparing cross-national non-response processes. Section 6 discusses limitations that researchers have encountered using currently available cross-national contact data.

Cross-national contact data: ESS, SHARE and ECHP

Comparative contact data from cross-national surveys are still rare. In the following, we introduce three studies where contact data were collected and have been analysed for response research. Research based upon these data is summarised in section 5.

The *European Social Survey* (ESS) was the first cross-national survey to collect and make publicly available cross-national contact data for both respondents and non-respondents. Except for a few countries, where data protection laws forbid the publication of all or parts of these data, the data of all countries are stored by the ESS data archive and can be accessed via its website (http://ess.nsd.uib.no/). The majority of countries follows the ESS recommendation of using the ESS model contact form. However, countries may use their own contact forms, provided that they submit all the compulsory variables to the data archive. Mandatory data collected include date, time, mode and detailed outcome of each contact attempt, interviewer number, reasons for refusal, whether a contact attempt is a re-issue and

whether the phone number of the sample unit is known. In addition to contact data, the ESS collects information on the housing and neighbourhood of the sample unit. Unfortunately, since many countries implement the contact forms on paper and do not carry out edit checks on the data, in some countries the data have many item missings.

The *Survey of Health Ageing and Retirement in Europe* (SHARE) collects and centrally stores call-level contact data through a common computerised case management system (CMS) for all sample units. In the first wave, 8 of the 11 countries participating in SHARE used the CMS. It allows interviewers to record the history of all contact attempts made to a household including the mode, the date, the time and the result code of each contact attempt. During interviewer training, which is streamlined across countries by means of a train-the-trainer system, interviewers are instructed on operating the system and on how to assign result codes. Unfortunately, the SHARE contact data are not publicly available and therefore not easily accessible for external methods researchers.

Data from the *European Community Household Panel* (ECHP) are available through the EuroPanel Users Network, epunet (http://epunet.essex.ac.uk). Only case-level contact data are available. For each sample household, for each wave of the survey, the relevant items available are the total number of visits made to the household, an indicator of whether the household was successfully traced, and indicator of whether the household was successfully interviewed.

5. Cross-national comparisons of non-response

In this section we summarise the methods and findings from cross-national comparisons of contact rates and response rates and possible determinants of response behaviour. This is a relatively young field of study. It effectively began in the 1990s with studies using country-level data on response rates. Only in recent years have a number of studies used case-level or call-level contact data.

Lyberg and Dean (1992) first coined the notion of different survey climates across countries and changes in survey climate within countries that affect response rates. Although only based on anecdotal evidence from the Netherlands, Germany and Sweden, the notion of survey climates has remained an important concept until today. Groves and Couper (1998) called for assembling time trends in response rates across countries. Supplemented with meta-data on “social environmental correlates of survey participation (e.g. degree of urbanization, level of political participation, alienation, education, crime rates, etc.), a

database could be built to permit comparative analysis of response rates within and across countries" (p.173). These kind of data they hoped "could reveal important differences in response to comparable surveys across countries that could be explained by variation in survey-taking climate" (p.172/3). De Heer (1999) published the first comprehensive study of cross-national non-response. Comparing surveys run by national statistical institutes in 16 countries, he found that there were "large differences in response rates and response trends between countries for official statistics." (p.140). This finding was supported by Couper and de Leeuw (2003), who examined response rates of three cross-national surveys (ISSP, IALS and TIMMS). The authors concluded that the differences in response rates and trends across countries and surveys indicated "differences in survey design and effort as well as societal differences" (p.165). In a more detailed multi-level analysis of de Heer's (1999) data, de Leeuw and de Heer (2002) found that survey-management and socio-economic meta-data on the survey and country level were predictive of country differences in outcomes. They determined factors that influence the non-contact rate (average household size, % of young children, panel surveys, lenient/strict rules for sampling and respondent selection) and factors that influence the refusal rate (% unemployed, inflation rate, mandatory surveys) (p.52/3).

While these studies granted important insights into cross-national *response rates* and *response trends* they inherit two main problems. First, they rely on macro-level data. As a result, they cannot make inferences about *response processes* across countries. De Leeuw and de Heer do find societal correlates of refusal and non-contact. However since all their data are on a macro-level (either on the survey or on the country-level), the analyses are vulnerable to an ecological fallacy (c.f. Robinson 1950). The macro-level finding for example that across countries household size is associated with non-contact, therefore does not necessarily imply that sample persons from large households across countries are more likely to be contacted. For individual-level inference one needs to have individual-level data. As a consequence, for analyses of the determinants of response and response bias one needs to analyse individual-level response data. Secondly, neither the surveys nor the calculations of response outcomes are necessarily comparable across countries. The studies rely on equivalent reporting over which they have no control. As a result, inferences about differences between countries are limited. As de Heer (1999) pointed out "without a detailed description of the response, it is impossible to evaluate the quality of a survey. Without comparable response rates it is difficult, to say the least, to compare or integrate data from different sources or countries." (p.141)

Studies using micro-level response outcomes and contact data fall into two groups. The first aim to explain differences in contact rates across countries. The underlying question is whether optimal contact strategies, defined as the most efficient number and timing of calls, are country specific or whether previous results from the US and UK (e.g. Bennett and Steel 2000, Campanelli et al 1997, Swires-Hennessy and Drake 1992, Weeks et al 1987) also apply to other countries. A related question is whether differences are explained by differences in the composition of national populations and of fieldwork characteristics, or by differences in response behaviours between countries. The answers to these questions have implications for how best to achieve equivalent fieldwork outcomes across countries: whether by standardising procedures, for example by requiring a specific minimum number of calls at specific times and days, or by allowing each country to adapt strategies to the situation it faces. The second group of studies examine to what extent contact data can help in identifying non-response bias. The studies test for correlations between information derived from contact data and both response propensities and survey variables. The findings have implications for the impact of non-response bias on cross-national comparative analyses and for methods of adjusting for non-response. All studies reviewed here use data from either round 1 of the ESS (Billiet et al. 2007; Blom, Jäckle and Lynn 2007; Kreuter and Kohler 2007; Kreuter et al 2007; Philippens et al. 2003), wave 1 of SHARE (Lipps and Benson 2005) or the 1995/6 waves of the ECHP (Nicoletti and Buck 2004).

The studies have firstly shown that there is considerable variation between countries in *fieldwork contact attempt procedures*. Philippens et al. (2003), for example, documented that some ESS countries had high proportions of non-contacted sample persons, who were called fewer than the required minimum of 4 attempts. The timing of calls, whether during a weekday daytime, evening or weekend, also varied hugely across countries. Nicoletti and Buck (2004) found that the distribution of fieldwork characteristics, conditional on individual and household characteristics, was slightly more similar between the independent BHPS and SOEP surveys (in Britain and Germany, respectively), than between the ECHP surveys in the same two countries, suggesting that the attempted harmonisation of fieldwork procedures in the ECHP was not successful.

Conclusions about whether *response behaviour*, that is the association between fieldwork procedures or population characteristics and contact outcomes, differs across countries are however mixed.

Philippens et al. (2003), Nicoletti and Buck (2004) and Blom et al. (2007) found country differences in the contactability and optimal calling strategies for the general

population, while Lipps and Benson (2005) found no country effects for the population aged 50+. This suggests that contact processes for the general population differ more across countries than contact processes for the population aged 50+. The different conclusions may, however, also be the result of the different analysis methods used in the studies. Philippens et al. (2003) tested for differences between countries in the association between the probability of making contact at the current call (conditional on not having made contact previously) and the timing of the call and number of previous calls, using a multilevel discrete time hazard model. The multilevel structure allowed for country specific intercepts and coefficients for the explanatory variables. The results indicated significant country-level intercept variance, suggesting that there were significant differences in the probability of making contact at the first call on a weekday afternoon or morning (the reference category). The coefficients for the timing of calls (whether evening or weekend) also varied significantly across countries, suggesting that "the 'optimal' timing of calls was country-dependent and illustrate[s] the importance of tailoring field work strategies towards specific national contexts" (Philippens et al. 2003, p.9). The number of previous contact attempts was negatively associated with the conditional probability of success in some countries, but had no effect in other countries. In addition, the country-level intercepts and residuals for the number of previous calls are negatively correlated, suggesting that "in countries for which respondents were 'easy to contact' at the first call, the probability of contact decreases more strongly than for countries in which respondents are 'hard to contact' at the first call" (Philippens et al. 2003, p.10). In sum, the study showed that countries differed in the ease of making contact, in the benefits of focusing on evening or weekend calls, and of increasing the minimum number of calls. Lipps and Benson (2005) firstly used all call attempts to estimate a variance components model of the probability of contact, allowing for country, interviewer- and household-level random intercepts. They secondly used only first call attempts, to eliminate confounding effects of contact with multiple respondents within a household or multiple contacts with the same respondent. To model the probability of co-operation, they thirdly included only the first successful contact attempt. In all three specifications, the country-level variance was barely or not significant, while the interviewer-level variance was. The authors then successively introduced covariates to attempt to explain the interviewer-level variance. In addition to indicators of the timing of a call (workday, Saturday, Sunday, morning, afternoon, evening, night) as used by Philippens et al., they also included an indicator for the mode of contact (face-to-face versus telephone) and interviewer observations about the physical state of the sampled address, the state of the neighbourhood and the existence of barriers to access of the

housing. The models allowed for random coefficients at the interviewer level, but not at the country level. The interviewer effects were however not explained by the available variables, suggesting that differences in outcomes between interviewers were not explained by differences in their contact strategies or characteristics of the assigned sample points. Instead, explaining interviewer effects would possibly require additional auxiliary information about the interviewers' characteristics.

Nicoletti and Buck (2004) and Blom et al. (2007) also tested for differences in the associations between fieldwork and population characteristics and outcomes. In addition, these studies examined whether differences in the *distributions* of fieldwork variables and population characteristics explained differences in response outcomes between countries. Nicoletti and Buck (2004) suggest that differences in contact and co-operation rates between Britain and Germany were mainly due to differences in response behaviours (model coefficients) rather than differences in population characteristics and that data collection variables were more important than individual and household characteristics. Harmonising fieldwork procedures would therefore not necessarily produce comparable response outcomes across countries. "Even if the explanatory variables distribution were equal between two surveys running in two different countries, the contact and the co-operation rates would not be equal because of a different impact of the variables. In other words the ease of contact and the propensity to co-operate, every explanatory variable being equal, are different across surveys running in different countries" (Nicoletti and Buck 2004, p.14). The comparison of eleven countries by Blom et al. (2007) further suggested that the reasons for differences depend on the countries compared and vary between countries with register and household/address based sampling frames.

The second set of studies examined associations between contact data and both response and survey outcomes, which would be indicative of non-response bias and could inform weighting adjustments. The associations between contact information and survey variables tend to differ between countries, but are generally small. Billiet et al. (2007) tested for *associations* between survey variables and respondent co-operativeness and showed that differences in summed attitude scores were larger between co-operative and hard-to-convert respondents (identified by the number of attempts required to persuade them to participate after an initial refusal), than between co-operative and easy-to-convert respondents. After controlling for differences in sample composition, by adding background variables thought to predict the attitudinal score, the associations between attitude scores and co-operativeness were reduced but not removed. This suggests that non-response bias could be adjusted further

by including information about co-operativeness in addition to the standard socio-demographic background variables used in non-response weighting models. The authors also concluded that the non-response bias was likely to be small, since removing the indicator of co-operativeness from the predictive model neither affected the explained variance nor the size of coefficients of the substantive covariates. Overall, the authors concluded that "the relationship between the type of respondent (co-operative, reluctant) and the attitudinal and background variables was not all in the same direction in all countries. This needs further research and discussion because it creates a serious challenge to any scholar who believes there is a theory of nonresponse that applies cross-nationally".

Kreuter et al. (2007) tested whether interviewer observations were correlated with survey variables. In addition, the authors tested whether the interviewer observations were correlated with response outcomes and what the effect of using them to construct non-response weights would be. The results suggested that correlations were low and varied by interviewer observation item, survey item and country. Weighting hardly changed point estimates. For the countries analysed the patterns were very similar, but exploratory analyses of other countries had apparently showed that this was not necessarily the case. The auxiliary variables tested were interviewer observations about the type of housing (whether multi-unit), signs of litter and vandalism in the neighbourhood and whether the sampled address had an alarm system. The survey variables were indicators of social involvement, fear of crime, general health and activities in the home. The approach is novel in that contact data and interviewer observations have mainly been used to predict response probabilities. Correlations with survey variables have rarely been tested. Correspondingly, contact data and interviewer observations have rarely been used for non-response adjustment.

Kreuter and Kohler (2007) used contact data to derive contact sequences and test hypotheses that these might be correlated with contactability, co-operativeness, interviewer behaviour and fieldwork regulations in different countries. If this holds, the authors propose using contact sequences for non-response adjustment. They define a contact sequence as a series of calls, which may either lead to no contact, contact with someone other than the sample person, contact with the sample person but no interview, or an interview. The contact sequence is composed of elements (each call attempt) and episodes (a sub-sequence of calls with the same outcome). The results indicate that the number of contact attempts is correlated with indicators of time spent at home (labour force status, time spent watching television). Similarly, countries which allow calls on Sundays produce contact sequences with fewer no contact calls.

6. Conclusion

In this paper we have identified a number of important potential gains from using contact data to study non-response in a cross-national comparative study. At the same time, there are also limitations to what can be achieved with contact data and even more severe limitations with what has been achieved to date.

Studies to date have principally been restricted by the extent to which contact data are available to analysts. It is no coincidence that five of the seven studies that have used cross-national micro-level contact data (see section 5 above) relate to the European Social Survey as this is the first, and thus far only, cross-national survey to make full contact data readily accessible. And even the analysts of ESS data have been restricted by limitations of the data. Kreuter et al (2007) only considered countries with near-complete interviewer observation data, with the result that only three of the 21 ESS round 1 countries are included in their analysis. Blom et al (2007) excluded countries from their analyses due to missing data. The approach of Billiet et al (2007), relying on refusal conversion to indicate reluctance, suffers a natural limitation in the sense that sufficiently large samples of converted refusers are required. Billiet et al were only able to include five of the 21 countries in their analysis.

Both Blom et al (2007) and Philippens et al (2003) point out that missing variables hamper the interpretation of their findings. For example, Philippens et al speculate that differences between countries in optimal calling strategies may be due to differences in at-home patterns, but are unable to explore this further without at least aggregate indicators of household size, household structure, employment patterns and so on. Blom et al speculate that additional micro-level measures for both responding and non-responding cases – additional to the limited interviewer observation items – might provide more explanatory power. Informative sampling frames such as population registers may be able to provide such measures. The sequence analysis approach of Kreuter and Kohler (2007) is limited by the length of the contact sequences. Longer sequences might produce greater variation in the sequences, permitting greater discrimination in terms of outcomes. The ESS specifies a minimum of four contact attempts before accepting a non-contact; a higher minimum might result in more informative contact data (though it would of course also have fieldwork cost implications).

Aside from data limitations, our knowledge of the ways in which contact data can help to explain cross-national non-response variations and to inform both fieldwork and adjustment strategies is limited by the fact that very few studies have yet been carried out.

Additionally, the few existent studies have differed in their objectives and in the methods used, as described in section 5 above. It is not yet possible to say with any confidence exactly what the role of contact data should be.

So, one might conclude that there is a need for call-level contact data to be made available on more surveys and for efforts to be made to make the data more complete, so that analysts do not have to drop whole countries from their analysis. This alone might stimulate further research and replication across surveys and survey organisations. However, some limitations will inevitably remain. In particular, we have highlighted the importance of equivalence in the procedures used to collect and capture contact data. This is perhaps the one area where even the most enlightened surveys, such as ESS and SHARE, have made only very limited advances to date. Kreuter et al (2007) point out that interviewer observation items are somewhat subjective and that this makes them susceptible to systematic differences between countries in the way they are interpreted and administered. Such differences would undermine their use in comparative analysis. The same is true of contact data, though it should be easier to standardise the implementation to a greater extent as the underlying phenomena are inherently factual. Interviewers should be given standard definitions (of what constitutes a call attempt, of each possible call outcome, etc) and some training in the application of those definitions (e.g. using vignettes or role-playing). Fieldwork quality control and monitoring systems should include contact data within their remit.

With standardised contact data, standardised implementation and better access to contact data for researchers, remaining limitations will be restricted to those imposed by the survey process. Missing variable problems will continue to be more intractable in countries where sampling frames are uninformative, compared to countries able to use, and extract information from, population registers or other micro-level sources. Analysis approaches such as those of Billiet et al (2007) and Kreuter and Kohler (2007) will continue to be subject to unquantifiable differences between countries in fieldwork processes. In the case of Billiet et al, countries will differ in the extent to which they attempt to convert refusals and in the criteria that they use to determine whether a particular case should be subject to a conversion attempt. In the case of Kreuter and Kohler, countries will differ in their distribution of call attempts and in the criteria that interviewers or field managers use to determine whether a particular case warrants additional call attempts.

In section 4 above we identified four criteria for standards regarding survey contact data. We suggest that the development of, and adherence to, such standards is likely to improve the quality of cross-national research into survey non-response and thereby to extend

our knowledge of the factors driving differences between countries in survey outcomes. We have discussed here in section 6 some of the ways in which measures could be taken to move towards these standards and the limitations that these measures would overcome. We would also note that our criterion 3 (equivalence of coding outcomes) would become redundant if surveys were able to make available full call-level contact data that had been collected in a fully standardised way. However, we have also pointed out that some limitations will remain. Researchers will need to be aware of these and to consider other ways of addressing them. But in conclusion, we believe that the availability of high-quality comparative contact data is in sight. We encourage all those with influence over cross-national surveys to bring this about by encouraging developments in the directions that we have outlined.

References

American Association for Public Opinion Research. (2000). *Standard definitions – Final dispositions of case codes and outcome rates for surveys.* Ann Arbor, MI: AAPOR.

American Association for Public Opinion Research. (2006). *Standard definitions – Final dispositions of case codes and outcome rates for surveys. 4th edition.* Lenexa, Kansas: AAPOR.

Bennett, D.J. and Steel, D. (2000). An evaluation of a large-scale CATI household survey using random digit dialling. *Australian and New Zealand Journal of Statistics,* 42, 255-270.

Billiet, J., Philippens, M., Fitzgerald, R. and Stoop, I. (2007). Estimation of nonresponse bias in the European Social Survey: using information from reluctant respondents. *Journal of Official Statistics,* 23, 135-162.

Blom, A.G., Jäckle, A. and Lynn, P. (2007). Explaining differences in contact rates across countries. Presented at *European Survey Research Association Conference*. Prague.

Campanelli, P., Sturgis, P. and Purdon, S. (1997). *Can you hear me knocking: an investigation into the impact of interviewers on survey response rates*. London: National Centre for Social Research.

Council of American Survey Research Organizations. (1982). *Special report: on the definition of response rates.* Port Jefferson, NY: CASRO.

Couper, M. (2005). Technology trends in survey data collection. *Social Science Computer Review,* 23, 486-501.

Couper, M. and de Leeuw, E.D. (2003). Nonresponse in cross-cultural and cross-national surveys. In J. Harkness, F.J.R. van de Vijver and P. Mohler (Eds.), *Cross-Cultural Survey Methods.* New Jersey: John Wiley.

Couper, M. and Lyberg, L. (2005). The Use of Paradata in Survey Research. *Proceedings of the 55th Session of the International Statistical Institute* (CD-ROM).

de Heer, W. (1999). International response trends: results of an international survey. *Journal of Official Statistics,* 15, 129-142.

de Leeuw, E. and De Heer, W. (2002). Trends in household survey nonresponse: a longitudinal and international comparison. In R. Groves, D.A. Dillman, J.L. Eltinge and R.J.A. Little (Eds.), *Survey nonresponse.* New York: John Wiley.

Groves, R.M. and Couper, M.P. (1998). *Nonresponse in household interview surveys*. New York: John Wiley.

Groves, R.M., Singer, E. and Corning, A. (2000). Leverage-saliency theory of survey participation. Description and illustration. *Public Opinion Quarterly,* 64, 299-308.

Groves, R.M., Fowler, F.J., Couper, M.P., Lepkowski, J.M., Singer, E. and Tourangeau, R. (2004). *Survey methodology*. New York: John Wiley.

Groves, R.M. and Heeringa, S.G. (2006). Responsive design for household surveys: tools for actively controlling survey errors and costs. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 169, 439-457.

Kreuter, F. and Kohler, U. (2007). Analyzing sequences of contacts. Presented at *European Survey Research Association Conference*. Prague.

Kreuter, F., Lemay, M. and Casas-Cordero, C. (2007). Using proxy measures of survey outcomes in post-survey adjustments: examples from the European Social Survey (ESS) and potential of the Los Angeles Neighborhood Observations Study (LA FANS). *JSM proceedings.*

Lipps, O. and Benson, G. (2005). Interviewer effort and data quality: paradata analysis of a cross-national survey, *Working Paper 1_05*. Lausanne: Swiss Household Panel.

Little, R.J.A. and Rubin, D.B. (2002). *Statistical Analysis with Missing Data*, 2nd Edition. New York: John Wiley.

Lyberg, L. and Dean, P. (1992). Methods for reducing nonresponse rates: a review. Paper presented at the annual meeting of the *American Association for Public Opinion Research*. St. Petersburg, FL.

Lynn, P., Beerten, R., Laiho, J. and Martin, J. (2001). Recommended standard final outcome categories and standard definitions of response rate for social surveys. *ISER Working Paper 2001-23.* Colchester: University of Essex.

Lynn, P., Clarke, P., Martin, J. and Sturgis, P. (2002). The effects of extended interviewer efforts on nonresponse bias. In D.A. Dillman, J.D. Eltinge, R.M. Groves, R. Little (Eds.), *Survey nonresponse*. New York: John Wiley.

Lynn, P., Japec, L. and Lyberg, L. (2006). What's so special about cross-national surveys? *ZUMA Nachrichten Spezial*, 12, 7-20.

Börsch-Supan, A. and Jürges, H. (eds.) (2005). *The Survey of Health Aging, and Retirement in Europe – Methodology.* Mannheim: Research Institute for the Economics of Aging.

McCarty, C. (2003). Differences in response rates using most recent versus final dispositions in telephone surveys. *Public Opinion Quarterly*, 67, 396-406.

Nicoletti, C. and Buck, N.H. (2004). Explaining interviewee contact and co-operation in the British and German Household Panels. In M. Ehling and U. Rendtel (Eds.), *Harmonisation of panel surveys and data quality.* Wiesbaden: Federal Statistical Office.

Philippens, M., Loosveldt, G., Stoop, I. and Billiet, J. (2003). Noncontact rates and interviewer calling strategies in the ESS. Presented at *International Workshop on Household Nonresponse*. Leuven.

Robinson, W.S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review,* 15, 351-57.

Smith, T.W. (1978). In search of house effects: a comparison of responses to various questions by different survey organizations. *Public Opinion Quarterly,* 42, 443-463.

Smith, T.W. (1982). House effects and the reproducibility of survey measurements: a comparison of the 1980 GSS and the 1980 American National Election Study. *Public Opinion Quarterly,* 46, 54-68.

Swires-Hennessy, E. and Drake, M. (1992). The optimum time at which to conduct interviews. *Journal of the Market Research Society,* 34, 61-72.

Weeks, M.F., Kulka, R.A. and Pierson, S.A. (1987). Optimal call scheduling for a telephone survey. *Public Opinion Quarterly,* 44, 101-114.