ISER Working Paper Series

Measuring Nonresponse Cross-Nationally

Annelies G. Blom
GESIS - Leibniz Institute for the Social Sciences
ISER, University of Essex

No. 2008-41
November 2008

INSTITUTE FOR SOCIAL
& ECONOMIC RESEARCH

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

www.iser.essex.ac.uk

Non-technical summary

When inferring from survey data to the general population researchers assume that the data were collected from a probability sample and that each sampled unit (e.g. each person) was actually interviewed. However, in every survey some persons are not interviewed, i.e. they are nonrespondents. When comparing data across countries, cross-national differences in nonresponse should be considered. However, measuring nonresponse on cross-national surveys is not straightforward. Differences in the implementation of a survey and in the population studied bring about differences in the prevalence of different sources of nonresponse. In order to measure nonresponse cross-nationally, countries need to coordinate how they collect nonresponse outcomes when implementing their surveys.

This paper reflects on three aspects of measuring nonresponse on cross-national surveys. Section one looks at factors that influence the response outcomes available to nonresponse researchers. Two groups of influences are considered: (1) constraints posed by the survey design and population structure and (2) data specifications that survey managers or survey methodologists develop. Section two develops a codeframe for collecting comparative response outcomes. An interviewer might need to make several contact attempts to a person before this person is contacted and willing to be interviewed. Therefore the outcome of each contact attempt has to be considered when describing the nonresponse process. In addition, we need to determine what the final outcome for each person is. The codeframe of response outcomes therefore describes the outcomes of each contact attempt as well as an overall outcome of each person. Since countries differ in their population and survey implementation characteristics, the codeframe takes country-specific outcomes into account, while simultaneously aiming for cross-national comparability. Finally, the third section evaluates two strategies (priority coding and most-recent coding) for assigning a final case outcome for each person. Analysing data on the nonresponse process in the European Social Survey the section concludes that priority coding is better suited to describe the nonresponse process. More importantly, the coding strategy influences the magnitude of contact and cooperation rates reported for a survey. Therefore, if countries use different strategies, this affects the comparability of outcome rates across countries.

Measuring Nonresponse Cross-Nationally

Annelies G. Blom

November 2008

Abstract

This paper investigates the measurement of nonresponse outcomes across countries. It consists of three sections. Section one proposes a conceptual framework of influences on the response outcomes available for analysis. Section two develops a cross-national codeframe of response outcome codes. The last section investigates the impact that differential ways of deriving final case outcomes have on the estimated response, contact and cooperation rates. The results emphasise the importance of careful measurement of the nonresponse process for conclusions about the similarity and differences across countries in processes leading to nonresponse.

Keywords: contact data, response outcomes, disposition codes, comparative research, response rates, European Social Survey

JEL codes: C81, C83

Contact: Annelies Blom, GESIS - Leibniz Institute for the Social Sciences,
P.O. Box 122155, 68072 Mannheim, Germany. Email: annelies.blom@gesis.org

Acknowledgements
Part of this paper is based on work carried out during my 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. I am most grateful to Peter Lynn for guidance and encouragement, to Mick Couper for feedback on UK-centric terminology and to the participants in the 2008 International Workshop on Survey Household Nonresponse in Ljubljana for comments on section three.

Introduction

Nonresponse as a source of error in survey data has gained importance over the past decades as the magnitude of nonresponse in many countries has increased (de Heer, 1999). The concern about nonresponse in sample surveys stems from a concern about nonresponse bias in the survey data. Statistical inference assumes that the data were drawn from the population by means of probability sampling and that each sampled unit is actually interviewed and their data observed. In survey reality however there are always sample units that cannot be interviewed. If the data missing due to nonresponse are missing systematically, there will be nonresponse bias.

The use of paradata (i.e. data about the survey process) for analysing and adjusting for nonresponse has surged since the introduction of computer assisted survey information collection (CASIC). In CASIC surveys some paradata are generated as a by-product of the survey process and can be automatically collected (Couper, 1998). Contact data are a type of paradata and they describe certain elements of the data collection process. In fully automated settings, e.g. in web surveys and many telephone surveys, contact data are collected automatically. Moreover, surveys were the data collection process is not automated, e.g. many face-to-face surveys, might collect contact data.

Contact data currently play a central role in nonresponse in sample surveys. They have two general purposes: to look at nonresponse bias and to monitor and analyse fieldwork processes. Contact data are especially suitable for bias analyses, if they are available for respondents and nonrespondents and are related both to the survey process and to the substantive survey outcome (Kreuter, Lemay and Casas-Cordero, 2007). Survey organisations have long been using contact data to monitor fieldwork and case management is usually conducted by means of data on the data collection process (i.e. contact data). Achieving high response rates in a cost-effective manner has traditionally been the focus of fieldwork monitoring, though some survey organisations are now moving towards looking directly at nonresponse bias by means of responsive designs (Groves and Heeringa, 2006).

Despite the widespread use of contact data for nonresponse research and fieldwork monitoring, standards on collecting them are still sparse. Yet to compare nonresponse

studies across surveys standardisation of the indicators of nonresponse is indispensable. Indicators of nonresponse can exist at different levels (Carton and Blom, 2008). Arguably the most prevalent are indicators (1) at the survey level (i.e. different types of response outcome rates, such as the response rate, the contact rate and the cooperation rate), (2) at the level of the sample unit (i.e. case-level response outcomes, such as the case outcome also called the final disposition of a case) and (3) at the level or the contact attempt (i.e. call-level response outcomes, such as the call outcome). Some progress towards achieving standardised response outcomes and outcome rates has been made at the national level in the US (AAPOR, 2006) and the UK (Lynn et al., 2001). However, no guidelines for comparing response outcomes and outcome rates across countries exist to date.

This paper develops a conceptual framework of influences on call- and case-level response outcomes in sample surveys. It looks into what we need to standardise (and what we should not standardise) and how we can arrive at comparable response outcomes and response rates. It then proceeds to develop a codeframe of response outcome codes for cross-national surveys where the sample unit is a person within the household. Using this codeframe a high level of comparability of response outcomes can be achieved. Analysing response outcomes from the European Social Survey (ESS) the paper finally investigates whether standardisation matters when deriving case outcomes from call-level contact data.

Background

The concern about an absence of comparable response outcomes is not new. Already in 1977 Kviz noted that "[t]he absence of a standard definition [of response rates] has caused a great deal of confusion regarding the interpretation of reported response rates and has frustrated methodological investigations because of a lack of comparative data." (Kviz, 1977, p.265) Kviz's paper was a milestone in defining the difference between the response rate (i.e. the number of achieved interviews or completed questionnaires divided by the number of eligible sample units) and the completion rate (i.e. the number of achieved interviews or completed questionnaires divided by the sample size). Just a few years later in 1982 the Council of American Survey Research Organizations (CASRO) took the response rate definitions a step further and

published a report *On the Definition of Response Rates*, which put forward basic definitions of response rates and illustrations of their calculation (CASRO, 1982). Though both of these papers covered important new ground they lacked definitions of actual case outcome codes, which are necessary to calculate outcome rates (including response rates).

In the late 1990s progress was made on the development of case outcomes and response rate definitions, when an AAPOR committee "developed standard definitions for the final disposition of case codes and of various outcome rates (e.g., response rates and cooperation rates) based on these codes" (Smith, 2002, p.30). Since, AAPOR has regularly updated their case outcome and outcome rate definitions for scholars and survey managers to adopt in surveys in the US.

Developing standards for UK surveys Lynn et al. (2002) noted that the AAPOR definitions were "limited and [were] not directly applicable to other countries." Amongst other things this was due to the fact that "[t]he nature of the sampling methods and sampling frames used for many social surveys in Europe … raises issues that are not dealt with in the AAPOR document" (Lynn et al., 2002, p.63). Lynn et al. (2002) developed case outcome codes for the UK, with detailed descriptions of their meaning and implementation. Though they seek a wider application of their definitions than the UK only, this endeavour is beyond the scope of their paper. They conclude "that the development of generic standards regarding the definition and implementation of survey outcomes and the definition and presentation of outcome rates – with cross-national applicability – is an important challenge whose time has come."(Lynn et al., 2002, p.77)

For cross-national surveys, de Heer (1999) voices concern about a lack of standards for response rate calculations when analysing an international trend in decreasing response rates. While his research showed that it was possible to collect response rates from various countries and various surveys – including cross-national surveys such as the Labour Force Survey – de Heer also noted that "extreme care should be taken that the data are comparable. … 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." (de Heer, 1999, p.140/1)

Though there have been cross-national surveys that have attempted standardising definitions of response outcome codes and response rate calculations (e.g. the European Community Household Panel, the European Social Survey and the Survey for Health Ageing and Retirement in Europe) they lack conceptual rationale for the outcome codes provided. As a consequence, outcome codes and their definitions in these studies are still very much survey- and country-specific.

An additional problem with currently available codeframes of response outcomes developed by AAPOR (2006) and Lynn et al. (2001) is that they regard case-level outcomes only. However, many response outcomes from the contacting process in the field are initially recorded by the interviewer at the call-level, i.e. the interviewer writes down or simply makes a mental note of the outcome of each contact attempt to the sample unit. "[T]he recorded status and contact history is used, typically by the interviewer, to decide when and how to make future calls." (Blom, Lynn and Jäckle, 2008, p.6) This call-level information may then be summarised by the interviewer, the supervisor or the survey agency, into case-level information which reports a final disposition for each sample unit (i.e. case-level outcome). To arrive at comparable case outcomes, however, both the definitions of response outcomes at the call level and the process of summarising call-level information have to be comparable. In addition, interviewers need clear and standardised information on what constitutes a contact attempt. Only if the same events are recorded across countries, can we compare the fieldwork processes. Especially, in face-to-face surveys and in cases where no contact was achieved this can be difficult.

The importance of call-level contact information has been noted by survey methodologists. In order to compare outcome rates calculated via standardised call outcomes the European Social Survey (ESS) was the first cross-national survey to collect detailed call-level cross-national data on the fieldwork process and make these publicly available for analysis (Stoop et al., 2003). In trying to achieve better comparability across countries and surveys, researchers such as those of the ESS

increasingly demand call-level contact data to consistently derive response outcomes and calculate outcome rates across surveys and countries.

This succinct overview of the background to comparative response outcomes and response rates describes that some progress has been made over the past decades, yet it also highlights that we are still some way from collecting response outcomes comparatively. The paper contributes to this area by developing a conceptual framework of influences on response outcomes, by building a codeframe of response outcomes that can be implemented across different countries and adapted for different kinds of surveys, and by evaluating the impact of differences in strategies for deriving a final case outcome from a sequence of call-level outcomes.

This work follows on from the equivalence criteria for comparative contact data developed by Blom, Jäckle and Lynn (2008). They note that "even with complete and publicly available data, inferences about cross-national differences in nonresponse are conditional on the equivalence of contact data" (Blom, Jäckle, and Lynn, 2008 forthcoming) and propose three equivalence criteria: (1) equivalence of design of contact data forms, (2) equivalence of implementation and (3) equivalence of coding fieldwork outcomes. Equivalence of design entails an equivalent structure and content of contact forms across countries. This includes a common codeframe of call outcomes across countries and equivalent structure of the contact forms to prevent question wording, format and translation effects. Equivalence of implementation includes the mode of data collection and of collecting contact data, briefing of interviewers regarding the collection of response data and the rules and control procedures governing contact attempts. Finally, equivalence of coding fieldwork outcomes concerns the coding of call outcomes for each sample unit into a final case disposition code for the sample unit.

In section one of this paper I develop a conceptual framework of influences on recorded fieldwork outcomes. This section underpins all three equivalence criteria. Section two, which develops a comparative framework of fieldwork outcome codes, primarily builds on equivalence criterion one. It also has implications for criterion two. The final section of this paper considers the extent to which the comparability of case outcomes depends on the coding strategy chosen to derive case outcomes from

call outcomes. It therefore examines the third equivalence criterion of Blom, Jäckle and Lynn (2008).

SECTION 1: CONCEPTUAL FRAMEWORK

The conceptual framework is based on the idea that two types of influences determine the contact data available to the survey manager or survey methodologist: First, there are constraints such as those posed by the population under study and choices about the survey design and implementation. These influence the *de facto* response outcomes that can be collected and made available for fieldwork monitoring and research. Second, the available response outcomes are influenced by the information that the survey manager or methodologist specifies to be collected or made available. In the following I will consider influences on the response outcomes of face-to-face surveys only. Similar sets of influences might govern the response outcomes of other modes of survey. This is especially the case for other interviewer-mediated surveys, such as telephone surveys. However, only face-to-face surveys are explicitly covered by this conceptual framework.

Constraints on fieldwork outcomes

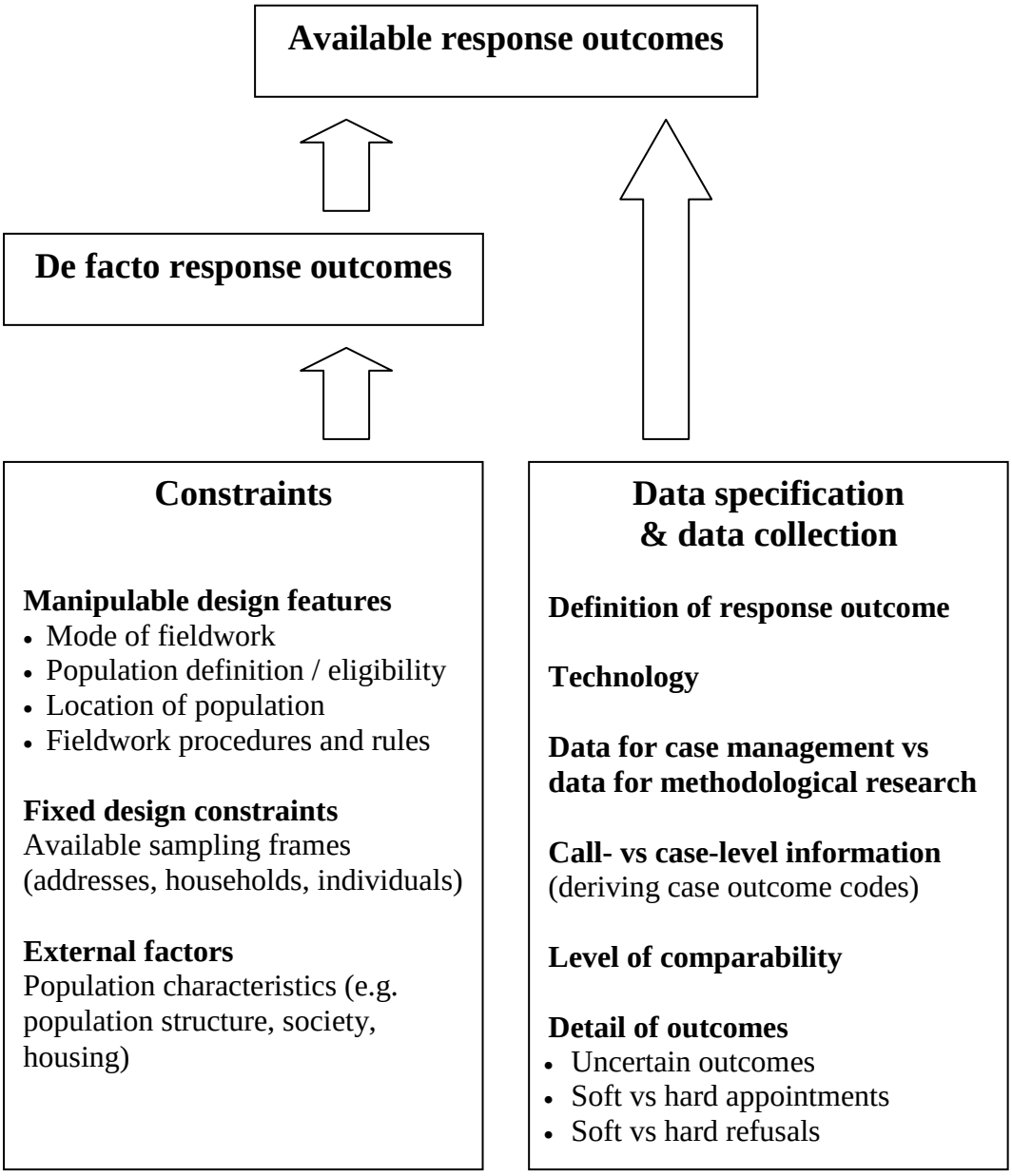
We can distinguish three types of constraints that determine the *de facto* response outcomes: (1) manipulable design features (i.e. design choices for which the researcher has multiple options from which to choose), (2) fixed design constraints (i.e. design features over which the researcher has no choice) and (3) external factors. Of these only the manipulable design features can be standardised to prevent differences in outcomes across countries. The fixed design constraints and external factors will differ across countries generating differences in the response outcome codes that need be specified and, as a consequence, also differences in the distributions of response outcomes.

Manipulable design features

The data collection process in face-to-face surveys is primarily carried out by means of in-person contact attempts by the interviewer. In addition other modes might be

permitted to make contact and gain cooperation. In most of Western Europe, for instance, survey organisations send advance letters before an interviewer attempts contact in person. If phone numbers are available for the sample unit, contact might even be attempted by phone before visiting the address.

Figure 1: Conceptual framework of influences on fieldwork outcome codes



Outcome codes might also need to be adapted to reflect the population studied. Many surveys in Europe for example exclude institutionalised persons (i.e. prisoners, people in care homes, people in the military etc) from the population researched. These people, if sampled, would then count as ineligible. Furthermore, many surveys restrict

themselves to certain age groups only. The ESS for example interviews persons aged 15 and over, while the Survey for Health Ageing and Retirement in Europe (SHARE) interviews those aged 50 and over. In cross-national surveys the country in which the survey is conducted plays an important role. The researcher needs to ensure that a sample unit has only one chance of being selected for participation in the survey. Therefore, special attention needs to be paid to sample units that move abroad, temporarily or long-term, and those who live in the country temporarily or long-term. Whether a unit is considered part of the population and therefore is eligible to be interviewed can also vary over time. In a sample of addresses someone might live at an address one day but have moved out the next. Trying to define what to do with cases where eligibility changes during the course of fieldwork, AAPOR (2006) coined the idea of defining a *status day*, "usually either the first day of the field period or the first day that a particular case was fielded" (AAPOR, 2006, p.18). If a sample unit is eligible on this specific day then it should be considered eligible for the survey. Unfortunately, this is impractical in actual field implementation, as interviewers often lack information to establish eligibility on status day, for example when a housing unit is found empty on the day that the interviewer visits. Furthermore, this means that if an interviewer has selected a person who is willing to do the interview, the interviewer would first need to make sure that this person was also eligible on status day. Similarly, in mail and web surveys with repeated invitations to the survey little is known about the eligibility on a specific status day. Defining a status day can therefore be an impractical and, for the interviewers, discouraging procedure. Even though with regards to strict probability sampling it is the correct procedure.

Furthermore, the setting in which the target sample units are contacted and interviewed shapes the outcomes that need to be taken into account. In surveys of pupils the target sample units are oftentimes all pupils of a certain grade, whereas in surveys of the general population sample units tend to be persons that are contacted in a household. The location of the target sample unit influences (a) the eligibility criteria of the sample unit and (b) the possible outcomes of the contacting and cooperation processes. In general population surveys that use household or address samples the interviewer typically selects one or more persons from the household for interview. The definition of what constitutes a household is therefore vital. "A country's operational definition of the private household is shaped by its national

culture. Each definition embodies a particular structure, and different definitions lead in turn to different structures with different compositions of the group definable as a household and, thus, to different household sizes." (Hoffmeyer-Zlotnik and Warner, 2008, p.11). The definition of a household may vary across countries (and frequently even does across surveys within one survey organisation). To a certain degree the definition of a household can be standardised across surveys. However, whilst in Western countries the definition of a household is comparatively straightforward[1], in cultures where the concepts of family and community differ from those of the West, defining a household can be tricky (cf. Quaglia, 2007; Lalor and Wardrop, 2007). In surveys of people within households the household definition plays an important role in defining outcomes of the contacting and cooperation processes. The reason is that contact with persons that do not belong to the household are counted as non-contact, "since these are not representatives of the targeted housing unit" (AAPOR, 2006, p.18). Contact in a fieldwork sense is therefore made as soon as the interviewer speaks to a household member. Contact with a neighbour, workmen in the house or a maid opening the door is considered a non-contact.

Finally, the fieldwork rules and procedures also influence the definition of *de facto* response outcomes. If refusal and non-contact cases can be re-issued, this should be reflected in the response outcome codes available (e.g. a 'converted refusal' code might be added). Similarly, in interviewer-administered surveys some organisations carry out back-checks on interviewers' work. This means that they ask the sample unit (in writing, by phone or in person) whether the interviewer actually visited them, carried out their work according to protocol and registered the correct response outcome. In some cases the back-checks find that the response outcome that the interviewer recorded is incorrect or that the interviewer committed fraud. In other cases, a back-check of a refusal leads to an interview, if the sample unit agrees to do the interview after all. The response outcomes collected need to reflect this possibility.

[1] Though there are many complexities at the margins, which surveys often ignore: e.g. children who divide their time between two parents, people living in two households (e.g. because the partner lives in a different place), etc

Fixed design constraints

Even in highly standardised surveys there are constraints on the implementation of the survey that cause differences in the possible outcomes, and hence in the required response outcome codes. A typical fixed design constraint in cross-national field implementation is the availability of suitable sampling frames. In surveys of individuals in households there are three general types of sampling frames: (1) lists of individuals (e.g. population registers), (2) lists of households and (3) address-based sampling frames (including random-route samples). "[C]ountries 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." (Blom, Lynn, and Jäckle, 2008, p.13) Since the availability of sampling frames is country-specific, response outcome codes that aim for cross-national comparability need to account for differences in sampling frames.

External factors

Differences in response outcome codes might also be necessary as a result of differences in characteristics of the population living in the countries under research. These external factors for example include national cultures in education, care for the elderly, military service or whether or how women can be interviewed. If a large proportion of the surveyed population is in the army, outcome categories would need to reflect this (cf. Lynn, 2003, p. 325). Similarly differences in housing might affect the response outcome codes necessary in a country. In some countries (e.g. the US or South Africa) gated communities are very common and the call outcome codes need to include something like 'no access to the address – gated community'. Similarly in many former Soviet countries blocks of flats do not always have doorbells at the entrance; instead an access code is needed to enter the building. An additional outcome code in such a situation might be 'no access to the address – housing block with access code'. For a cross-national team coordinating field implementation it is impossible to foresee all external factors that need to be accounted for when designing a codeframe of cross-national outcome codes. Communication about this with national experts is therefore essential.

As a final point, many of these constraints on the de facto fieldwork outcomes interact with each other. For instance, household definitions are less important if the sampling frame consists of individuals, since no selection within the household needs to be performed. External factors might also bring about procedural differences across countries. For example, the speed and reliability of the postal system influences organisations' decisions to send out advance letters. Furthermore, legal constraints influence fieldwork procedures. In some countries re-attempting refusals is illegal and this also influences back-checking procedures. Interactions are also found between constraints above and the data specifications described in the following.

Data specification and data collection

As described above, the *de facto* response outcomes that are obtained in the field depend on constraints posed by the survey design and the population studied. A codeframe of response outcome codes needs to account for these influences. In addition, the survey manager or methodologist can specify response outcomes that they want the interviewers to record; they can specify response outcome codes for fieldwork monitoring in case of the survey manager and codes for nonresponse research in case of the methodologist. Contact data designed by the survey manager for case management, i.e. to monitor progress of each case during fieldwork, might lack information that the survey methodologist needs for her analyses and vice versa. Traditionally, contact data were designed for monitoring fieldwork, such as checking interviewer performance, controlling costs and achieving high response rates. This means that some contact data are less useful as a data source for nonresponse research.

Definition of response outcome

The definition of what constitutes a response outcome influences which *de facto* response outcomes we collect. Usually a response outcome is collected at a contact attempt. However, what is a contact attempt? In telephone surveys each phone call is usually regarded as a contact attempt, however, in face-to-face surveys issues are less straightforward. Does driving by a house to check out the area (without ringing the doorbell) count as a contact attempt? Probably not. However, setting out to visit a sample unit, but not being able to reach the house (i.e. not ringing the doorbell) probably would. Furthermore, there are response outcomes that do not stem from the

field, so-called in-office outcomes, that are collected in the response outcomes. Examples include office refusals (i.e. cases where the sample unit contacted the survey organisation directly to refuse participation in the survey) which can take place before or after the interviewer has conducted contact attempts. Also rejections of interviews at the editing stage or downgrading from full to partial interview might be recorded in the response outcomes. To have a full record of the nonresponse process these kinds of outcomes should be explicitly recorded, perhaps as an extra call in the contact data where the interviewer identifier is set to 'office'.

A related issue is that some survey organisations collect activity codes in addition to response outcomes. These activity codes register fieldwork efforts that have been undertaken, but not necessarily by the interviewer. The sending of advance or reminder letters, payment of incentives, giving the sample unit a self-completion questionnaire that needs to be sent back to the office and back-checking efforts are examples of information recorded in such activity codes. Activity codes are no response outcomes, yet they describe to the fieldwork process.

Technology for collecting response outcomes

The technology with which contact attempts are recorded can influence response outcomes collected during the fieldwork process. While telephone surveys usually record each contact attempt automatically in a computer system, face-to-face surveys differ in their recording of contact attempts across survey organisations. In some organisations interviewers record their efforts on paper forms, while others use computer systems (often the same systems that also record the interview). Furthermore, even if computer systems are used to record the contacting process, the organisation might provide interviewers with additional paper forms for noting down contact attempts whilst working in the field. Only at the end of their working day do interviewers complete their electronic contact forms (Couper, 2008). These differences in procedures can lead to differences in the types and quantity of contact attempts recorded.

Case management systems and data for methodological research

How data are captured depends, in part, on how they are used. How contact data are used also determines whether contact data are recorded as call-level data (i.e. at the

level of each contact attempt) or summarised and recorded as case-level data (i.e. at the level of the sample unit). For case management, face-to-face interviewers may use full call-level data about each past call 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. 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, though summary measures (such as number of call attempts) might be. Alternatively, the case management system may not record the outcome of past calls at all, but instead record the current status of a case (i.e. an intermediate case outcome) and what the next action should be. These systems 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. Survey methodologists often prefer full call-level information, where the interviewer records and transmits every contact attempt. The researcher is then free to derive a several case-level summary measures from the call-level data, such as total number of calls, timing of calls, intermediate outcomes, interviewer contact and cooperation rates etc. Recently, analyses of contact data and field interventions have been tested in real-time systems, so-called responsive designs. 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 fieldwork 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 final case outcomes (e.g. interview, refusal, non-contact, ineligible) and the total number of calls/visits made. A special case of interviewers summarising call outcomes to case outcomes is where the survey organisation works with issue outcomes. An issue outcome is the outcome for a sample unit that has been achieved by an interviewer and returned to the office. To increase response many survey organisations re-examine the outcome of each unproductive sample unit in the office and decide on whether to re-issue the sample unit to the field, either to the same or to a different interviewer. At the end of fieldwork a sample unit can therefore have more than one issue outcome. In the second case survey organisations do receive call-level contact data from the interviewers, but they might decide to derive final case outcomes from the call outcomes before passing the data on to a third party. 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, methodologists may derive final outcome codes at the case level from call-level response outcomes in order to use their own preferred definitions in analysis.

Irrespective of the stage at which the case outcome is derived, there are three main methods by which it may be derived: *most-recent*, *priority* and *subjective* coding. With most-recent coding, the outcome of the last call to a sample unit is designated as the case outcome AAPOR, 2000). Priority coding, on the other hand, involves arranging call outcomes according to a priority ranking, 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 (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 defined 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 deriving a case outcome from call outcomes are not defined. Typically in such situations, only descriptions of each case outcome code (which can vary in their precision) 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.

Level of comparability

When specifying which response outcomes codes are to be provided to the interviewer, the researcher or survey manager decides on how detailed the information will be. The amount of detail can be decisive when considering whether response outcomes across surveys are comparable. In section 2 below I develop a case-level codeframe of response outcome codes, where I implement a three-level hierarchy. At the highest level response outcomes are comparable across countries and types of surveys, since only very general response outcome codes are specified: ineligible, unknown if eligible, non-contact, contact with household (but not refusal or interview), refusal and interview. If only this level of detail is collected in a survey, analyses of the data and comparisons across countries would be severely limited.

Detail of response outcomes

If response outcomes are registered too crudely, this can be problematic for monitoring fieldwork progress, carrying out quality checks and for conducting nonresponse analyses. While previous codeframes of response outcome codes (cf. Lynn et al., 2001; AAPOR, 2006 and the ESS outcome codes in section 3) are detailed and well-established, in some areas they lack outcome codes that are relevant for this. Codes for uncertainty of outcomes, soft and hard appointments and soft and hard refusals are often measured in little detail. They are discussed in the following.

Uncertainty of response outcomes

The degree of certainty of a response outcome is an important issue, both in managing fieldwork and for analysing fieldwork processes. The main category of response outcomes that is affected by uncertainty is that of eligibility. In any survey

interviewers will encounter cases where they cannot establish without doubt whether a case is eligible or not. Interviewers might not get access to the sample unit (e.g. in the case of gated communities), are refused the information necessary to ascertain eligibility or the sample unit is never issued to the field. For a full record of the survey process it is important to give the interviewer the possibility to register an uncertain response outcome. With response outcomes that indicate uncertainty the survey organisation can trace the status of a case and use the information provided by the interviewer to make decisions about how to clarify the response outcome of a sample unit (e.g. by sending a more experienced interviewer). If eligibility cannot be ascertained by the end of fieldwork the contact data will have a record of this. When researching processes leading to contact and cooperation and when looking at nonresponse bias, data on calls or cases with uncertain outcomes hold valuable information about the sample unit (cf. Brick, Montaquila and Scheuren, 2002). For example cases that could never be accessed and their eligibility was never established are likely to be different in key characteristics from cases that refused the information necessary to establish eligibility.

Lynn et al. (2001) include categories for unknown eligibility in their codeframe. In addition, they suggest that, when calculating a response rate, researchers should estimate the proportion of the sample units of unknown eligibility that were actually eligible. Furthermore they distinguish between cases of unknown eligibility, where (a) contact with the household has been made (e.g. when contact has been made but the household refuses to give the information necessary to establish whether someone fulfilling the eligibility criteria lives in the household) and (b) no contact has been established at all. AAPOR (2006) deals with this issue slightly differently. They also include codes for unknown eligibility in their codeframe. When calculating the response rate, they propose different kinds of rate calculations, of which some (RR1 and RR2) count all cases of unknown eligibility as eligible, while others (RR3 and RR4) include an estimate of the proportion of eligible versus ineligible sample units. However, they do not distinguish between unknown eligibility of contacted and non-contacted cases.

Yet, identifying the cause of an uncertain outcome (e.g. uncertainty due to non-contact, refusal or other nonresponse) is pivotal, because the processes leading to non-

contact and refusal are distinctly dissimilar (cf. Lynn and Clarke, 2002). Traditionally codeframes of fieldwork outcomes do not provide detailed data on uncertain outcomes and their cause. However, such information might add explanatory power to models of nonresponse processes and to the calculation of response rates.

How to deal with cases of uncertain eligibility when calculating response rates can be an intricate and political issue; especially where country comparisons are concerned. The proportion of cases of unknown eligibility that was actually eligible might differ across countries. To estimate these different proportions for each country necessitates in-depth knowledge of each country's fieldwork situation and population structure, which is seldom available at the level of cross-national coordination of a survey. If this estimation is undertaken at the national level on the other hand, there is scope for influencing the estimated response rate by counting a larger proportion of cases ineligible than reasonable. As Lynn et al. (2002) illustrate, in the example of the Welsh Assembly Election Study the estimated response rate ranged from 30.3% to 36.5% depending on the estimated proportion of eligible cases amongst those of uncertain eligibility. To complicate matters further the variety of response outcome codes for unknown eligibility can be large and the proportions of cases eligible will vary according to outcome code. In some surveys one can assume that almost all of the cases of an outcome code are actually eligible. In the ESS for example all persons aged 15 and older are eligible. Since it is unlikely that a household in Europe contains only persons under the age of 15, one can assume that close to 100% of resident households at ESS sample addresses contain at least one eligible person. For other uncertain-eligibility outcome codes the proportion of eligible cases is likely to be less than 100% and to vary according to code and country.

One approach to this issue is to count all cases of unknown eligibility as eligible. This leads to conservative response rate estimates and encourages interviewers and fieldwork staff to clarify as many cases are possible to increase the apparent response rate. An alternative way of dealing with cases of unknown eligibility is to calculate response rate boundaries. The upper boundary gives the response rate, if all cases of uncertain eligibility were ineligible (i.e. the maximum response rate), while the lower boundary gives the response rate, if all cases of uncertain eligibility were eligible (i.e. the minimum response rate). The true response rate lies within these boundaries.

RR_{LB}	<	RR	<	RR_{UB}
(all uncertain cases eligible)		true response rate		(all uncertain cases ineligible)

The more cases of uncertain eligibility there are amongst the final outcome codes, the larger the response rate interval will be. The response rate interval therefore gives an estimate of the response rate as well as an indication of how precise the estimate is.

In summary, whether and which outcome codes for uncertain outcomes are provided in a survey will depend on the population researched and, more importantly, what the survey managers and methodologists deem necessary for their work. It is crucial to note that the choice of outcome codes to represent uncertainty as well as the treatment of cases of uncertain eligibility will affect the comparability of the response data.

'Hard' and 'soft' appointments and refusals

Appointment and refusal outcomes are categories of outcome codes that tend to differ widely in detail across surveys. The call outcome 'appointment' and case outcome 'refusal' are closely related, because nonresponse researchers usually count broken appointments as refusals (cf. Billiet et al, 2007, p.142). This means that if a (broken) appointment is derived as the final case outcome, this is counted as an implicit refusal.

With regards to refusals many researchers distinguish 'hard' and 'soft' refusals. The definitions of what constitutes 'hard' and 'soft' can vary. Frequently, the reasons for refusing are consulted to distinguish between the two. A reason such as, being momentarily busy, would count as soft refusals, whereas hard refusal could be people who do not want to participate for ideological reasons. Another way to distinguish hard and soft refusals is by asking the interviewer to assess how likely sample units' participation is when re-attempting them (for an operationalisation of the 'softness' of refusals in the ESS consult Loosveldt et al, 2003). Consequently, soft refusals are those where a conversion attempt (possibly by a different interviewer) is likely to yield an interview. Hard refusals are usually not reattempted, because the sample person made clear that she is not going to participate in the survey. Distinguishing between hard and soft refusals can be useful when analysing response, for example by means of a classes of respondents approach (cf. Lin and Schaeffer, 1995; Stoop,

2005). However, for valid comparisons the definition of what constitutes a soft or hard refusal needs to be consistent across countries.

Unfortunately, this distinction between 'soft' and 'hard' is not generally applied to appointment outcomes (cf. AAPOR, 2006; Lynn et al., 2001). Yet when analysing fieldwork processes it might be important to distinguish between hard appointments, i.e. appointments made directly with the sample unit for a specific day and time, and 'soft' appointment, i.e. appointments made by another household member on behalf of the sample unit (e.g. a household member indicates to the interviewer, when would be a good time to reach the sample unit). A broken hard appointment is much more clearly an implicit refusal than a broken soft appointment. A broken soft appointment might also be considered a refusal; however, this refusal is less severe than when a hard appointment is broken. It could simply indicate that the person to whom the interviewer spoke was misinformed about when the sample member would be at home.

Distinguishing between hard and soft appointments and refusals should be especially fruitful in cross-national surveys, where differences in the population structure, social settings and fieldwork practices can lead to different proportions of hard and soft refusals and appointments.

SECTION 2: A CROSS-NATIONAL CODEFRAME OF RESPONSE OUTCOMES

The conceptual framework of influences on response outcome codes sets the scene for the following section, where I propose a cross-national codeframe of response outcomes. The aim of the codeframe developed in this paper is to yield response outcome codes that allow maximum comparability of fieldwork outcomes in cross-national comparisons. Consequently, the focus of the codeframe is comparability in *post hoc* analyses; either in analyses of nonresponse processes or analyses of nonresponse bias. The case management aspect of fieldwork outcomes is however not ignored. The structure of the codeframe and its discussion take implementation practicalities into account.

A combined call and case-level codeframe

The codeframe is a combined call- and case-level frame. This combination of call- and case level outcomes stems from the nature of field outcomes and from how a final case outcome for a sample is established. We can distinguish three types of response outcomes (Figure 2): (1) *In-office case outcomes* are response outcomes that do not originate from the field process, but instead are assigned to a sample unit in the office. In office case outcomes might be either assigned to cases that have never been in the field or to cases that have been worked on by an interviewer, but where the final field outcome was superseded by an in-office case outcome. (2) *Final call outcomes* are call outcomes from the field, which will by their very nature equal the final case outcome. By definition there can only be one call outcome of this type for a case. The most obvious example for a final call outcome is when a full interview is achieved. This outcome is by its very nature the final call outcome of the case. (3) Lastly, *non-final call outcomes* do not have an automatic case outcome attached. Sample units that have a sequence of only non-final call outcomes do not have an obvious final case outcome. From the sequence of non-final call outcomes a final case outcome needs to be derived, i.e. a *derived final case outcome* needs to be assigned.

In-office case outcomes include sample units that were never issued to an interviewer and therefore only have a case outcome that was assigned to the sample unit in the office. Likewise an office refusal will typically, as a case outcome, take priority over any call outcomes from the field and is therefore an in-office case outcome.

Figure 2: Deriving case-level outcomes from contact data

Call-level outcomes	**Case-level outcomes**
(may or may not have call outcomes)	In-office case outcome (supersedes any call outcome)
Final call outcome (with or without non-final call outcomes) →	Final case outcome
Non-final call outcomes (only) →	Derived final case outcome

Where a final call outcome is achieved in the field, the interviewer will not return to the sample unit, because the achieved outcome renders this unnecessary or impossible. Most ineligibility outcomes are of this type. The contact attempt that establishes the ineligibility of a case is automatically also the last call attempt and

determines the final case outcome. Only if the survey organisation decides to re-issue a case to confirm or reject the ineligibility, might this code still be overwritten. Other call outcomes that can be considered final call outcomes include ‘hard’ refusals, after which the interviewer will not return to the sample unit again. (In some countries re-approaching ‘hard’ refusals is in fact illegal.) Completed interviews are the ultimate type of final call outcome, since the goal of fieldwork is achieved on the call at which the interview is completed.

In the case of a sequence of only non-final call outcomes the final case outcome is not directly clear and needs to be derived from the sequence of contact attempts made by the interviewer. An example of this is the common situation in which an interviewer re-attempts a sample unit several times, yet the unit is either repeatedly busy or not reached. The final case outcome will then need to be derived from the sequence of these non-final call outcomes. There is one exception to this typology, when a non-final call outcome is achieved through the office. This is might for example happen if a sample unit calls the operations department of a survey organisation to arrange an appointment for interview. This should be also be registered on the contact form as an ‘hard’ or ‘soft’ appointment with the sample unit. I consider this case a non-final call outcome.

Some fieldwork response outcome codes stem from the call level (final and non-final call outcomes) while others are case-level outcomes (in-office case outcomes). For many nonresponse analyses and for calculating response rates we need one outcome per case, i.e. case-level outcomes. As a consequence, the comparative codeframe of response outcome codes needs to account for this and combine the call and the case-level. Final call outcomes at the call level of the codeframe should have a direct match in the case level of the codeframe. Sequences of solely non-final call outcomes from the call level will need to be matched into derived final case outcomes at the case level.

A three-level hierarchy

The combined call and case-level codeframe aims to be sufficiently generic for a wide range of settings. At the same time cross-national comparability needs to be ensured. Nevertheless, the codeframe has to be practical and provide enough detail for field

implementation. Following to some extent Lynn et al. (2001) the codeframe is arranged as a three-level hierarchy. Level 1 can be implemented across countries and different types of surveys. The detail provided in level 2 restricts implementation to surveys of the same type. In this codeframe these are face-to-face surveys of individuals in households, where one individual per household is interviewed. Across countries, however, the same level-2 categories should be used for surveys of a given type. Level 3 can be adapted to fit differences across countries, such as population or housing structure. The codeframe can be adapted at level 3 to account for country-specific external factors and design constraints. The codeframe in this paper gives some examples of level-3 codes that might be relevant in some countries. However, acknowledging that detailed experience of society and fieldwork across a great range of different countries would be necessary to suggest further country-level adaptations the detail provided at level 3 can only be a starting point. For cross-national surveys of individuals (within households) I suggest six outcome codes at level 1:

1 Ineligible
The sample unit was selected from the sampling frame. However, it can be ascertained that the sample unit is not part of the population under study.

2 Unknown if eligible
The sample unit was selected from the sampling frame. However, it cannot be ascertained whether the sample unit is or is not part of the population under study.

Contacted and non-contacted sample units

3 Non-contact
The sample unit is eligible, but has never been contacted.

4 Contact with household, but no refusal or interview
The sample unit is eligible. However, only the household that the sample unit is located in is contacted. In the case of address and household sampling frames, the sample person might or might not have been selected.

5 Refusal
The sample unit is eligible. In the case of address and household sampling frames, the sample person might or might not have been selected. The sample person or the household refuses to participate.

6 Interview
The sample unit is eligible. In the case of address and household sampling frames, the sample person has been selected. The sample person has been interviewed.

Known to be eligible sample units

The hierarchical structure means that response outcomes of surveys that share the same outcome codes down to the third level will be better suited for comparisons than those that only share level-1 outcome codes.

The codeframe of case outcomes

The codeframe of response outcomes has two parts: a codeframe of case outcomes containing in-office case outcomes, final call outcomes and derived case outcomes (which were derived from a sequence of non-final call outcomes) and a codeframe of final and non-final call outcomes that can be coded into case outcomes. I first consider the codeframe of case outcomes (Table 1).

These codeframes are designed to closely match the fieldwork process. Following the codes in the codeframe from top to bottom the reader also follows the typical way an interviewer handles a case. The interviewer first establishes the eligibility of a sample unit (codes for ineligibility and uncertain eligibility). Subsequently the interviewer will attempt to make contact, but might not succeed right away (non-contact codes). Next contact at the household and then with the sample person might be established and an appointment might be made (and then broken) or a refusal is received (general contact and refusal codes). Finally, the interviewer manages to do the interview with the sample unit (interview codes).

The first column of the table below indicates the number and level of hierarchy of the case outcome; the second column describes the case outcome; the third column indicates if the outcome is restricted to a certain type of sampling frame (i.e. frames of individuals (IN), households (HH) or addresses (AD)); finally, the fourth column indicates whether the case outcome is an in-office case outcome (IO), a final call outcome (FCO) or whether the case outcome was derived from a sequence of non-final call outcomes (DFO).

Those familiar with codeframes of case outcomes, such as the codeframes of AAPOR (2006) and Lynn et al. (2001), will notice that some outcomes were adopted from these standardised codeframes while others are new. The new aspects of this codeframe especially concern the larger number of outcomes of uncertain eligibility and broken appointments. The rationale for these codes was discussed in the conceptual framework of influences on outcome codes in section one.

Table 1: Codeframe of case outcome codes
(for in-person surveys of one individual interviewed per household)

	Case outcome	**Sampling frame** Only sample type AD = addresses HH = households IN = individuals	**Type of outcome** IO (in-office case outcome) FCO (final call outcome) DFO (final outcome derived from sequence of call outcomes)
1	**Ineligible**		
1.1	Address not eligible	HH/AD	FCO / IO
1.1.1	Not yet built/under construction	HH/AD	FCO
1.1.2	Demolished/derelict	HH/AD	FCO
1.1.3	Vacant/empty	HH/AD	FCO
1.1.4	Non-residential address (business, hospital, school etc)	HH/AD	FCO
1.1.5	Address out of sample	AD	FCO / IO
1.1.6	Other	HH/AD	
1.2	Address occupied, but no eligible household	HH/AD	FCO
1.2.1	Vacation homes	HH/AD	FCO
1.2.2	Temporary residences	HH/AD	FCO
1.2.3	Other	HH/AD	FCO
1.3	Household occupied, but no eligible person		FCO
1.3.1	Nobody of eligible age in household	HH/AD	FCO
1.3.2	Sample person not eligible (e.g. not aged 18, moved abroad)	IN	FCO
1.3.3	Other		FCO
1.4	Other ineligible		FCO
2	**Uncertain eligibility**		
2.1	Never issued to an interviewer		IO
2.2	Issued but never attempted – unable to locate address/household		IO
2.3	Issued but never attempted – address/household inaccessible, eligibility unclear		IO
2.3.1	Unsafe area		IO
2.3.2	Bad weather conditions		IO
2.3.3	Gated community		IO
2.3.4	Housing block with access code		IO
2.3.5	Other		IO
2.4	Issued but not attempted – other		IO
2.5	Unknown whether address contains eligible household(s) – information refused	AD	DFO
2.6	Unknown whether address contains eligible household(s) – no information due to non-contact	AD	DFO
2.7	Unknown if selected household is eligible – information refused	HH	DFO
2.8	Unknown if selected household is eligible – information unavailable (e.g. language problems)	HH	DFO
2.9	Unknown if selected household is eligible – no information due to non-contact	HH	DFO
2.10	Eligible household, unknown if eligible	HH/AD	DFO

	person(s) – information refused		
2.11	Eligible household, unknown if eligible person(s) – information unavailable (e.g. language problems)	HH/AD	DFO
2.12	Eligible household, unknown if eligible person(s) – no information due to non-contact	HH/AD	DFO
2.13	Unknown if selected person is eligible – information refused	IN	DFO
2.14	Unknown if selected person is eligible – information unavailable (e.g. language problems)	IN	DFO
2.15	Unknown if selected person is eligible – no information due to non-contact	IN	DFO
2.16	Selected unit moved, not re-attempted	IN	DFO / IO
2.17	Other unknown eligibility		DFO / IO
3	**Non-contact**		
3.1	No contact made with anyone at the address		DFO
3.1.1	Household selection not achieved (e.g. gated community, housing block with access code etc), but certainty about eligibility	AD	DFO
3.2	Contact made at the address but access to household denied (e.g. concierge)		DFO
3.2.1	Information for household selection refused, but certainty about eligibility	AD	DFO
3.3	No contact with the sampled household		DFO
3.3.1	No selection of sample person, but certainty about eligibility	HH/AD	DFO
3.4	Contact made at the sampled household, but not with any responsible resident (e.g. visitor, workmen, maid, children)		DFO
3.4.1	No selection of sample person, but certainty about eligibility	HH/AD	DFO
3.4.2	Sample person moved within country, never contacted, but certainty about eligibility	IN	IO
4	**Contact with household (excluding refusal and interview)**		
4.1	Contact made at the household, but selection of person not achieved	HH/AD	DFO
4.1.1	Language problems at the household	HH/AD	DFO
4.1.2	Never a good time to call / always busy	HH/AD	DFO
4.2	Contact made at the household, but no contact with selected person		DFO
4.2.1	Never a good time to call / always busy		DFO
4.2.2	Sample person always away		DFO
4.3	Contact made with the selected sample person, but sample person unable to do the interview		FCO
4.3.1	Language problems with the sample person		FCO
4.3.2	Sample person physically or mentally unable to do the interview		FCO
4.3.3	Never a good time to call / always		FCO

	busy		
4.4	Other contact		
5	**Refusal**		
5.1	Office refusal before interview		IO
5.2	Contact made at the household with responsible resident – household refusal		FCO / DFO
5.2.1	Selection of sample person not achieved	HH/AD	FCO / DFO
5.3	Broken appointment; appointment made by household member on behalf of sample person		DFO
5.4	Proxy refusal by household member on behalf of the sample person		FCO / DFO
5.5	Broken appointment; appointment made by sample person		DFO
5.6	Refusal by sample person		FCO / DFO
5.7	Refusal during interview		FCO
5.8	Broken-off interview, never completed		DFO
5.9	Office refusal after interview – request for data to be deleted		IO
6	**Interview**		
6.1	Interview with sample person		FCO / IO
6.1.1	Partial interview with sample person		FCO / IO
6.1.2	Interview partly with sample person partly with proxy		FCO / IO
6.1.3	Interview with proxy		FCO / IO

Call-outcome codes in field implementation

When collecting response outcomes during fieldwork, only call-level outcome codes are implemented, i.e. the final call outcomes and the non-final call outcomes. These need to be collected by the interviewer and must be available to them. The interviewer assigns a call outcome to each contact attempt she makes. The case outcomes are then determined by these call outcomes from the field and the superseding in-office case outcomes. The codeframe of case outcomes described in Table 1 is therefore never directly implemented in the field, i.e. is not handed to interviewers. The derived final case outcomes in the case-level codeframe are derived from non-final call outcomes and the in-office outcomes are collected at the operations department. Only final and non-final call outcomes are directly collected during fieldwork on so-called contact forms.

Figure 3 shows an example of what a paper version of a contact form may look like. In this contact form the interviewer writes down the relevant outcome code for the respective call in the column ‘outcome of visit’. Some of these outcomes are final call outcomes, while others are non-final call outcomes from which final case outcomes

are derived. Contact forms might also be implemented on a computer. In this case, contact forms might be organised hierarchically. For each contact attempt the interviewer fills in a short questionnaire-like form, which includes routing based on information from previous contact attempts.

Looking back at the case-level codeframe, the observant reader will notice that many of the derived final case outcomes in the codeframe are a combination of the call-level outcomes and of whether the interviewer was able to select the household and sample unit. For household and address samples the contact form should therefore include a column that registers whether the household/sample unit was selected during a respective call, as Figure 3 illustrates. In computerised contact forms this information can be asked in the routed questionnaire.

Figure 3: 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)	Was household selected?	Was sample unit selected?	Notes
1	/	:			Y/N	Y/N	
2	/	:			Y/N	Y/N	
3	/	:			Y/N	Y/N	
4	/	:			Y/N	Y/N	
5	/	:			Y/N	Y/N	
6	/	:			Y/N	Y/N	
7	/	:			Y/N	Y/N	
8	/	:			Y/N	Y/N	
9	/	:			Y/N	Y/N	
10	/	:			Y/N	Y/N	

The call-level codeframe can thus be less complex than the case-level codeframe; in-office outcomes do not need to be taken into account and the selection of the household/sample unit is dealt with elsewhere on the contact form. Table 2 provides a list of call outcomes, from which, in combination with information on whether the household/sample unit was selected, the final case outcome can be derived.

Though more straightforward than the codeframe of case outcomes, some survey organisations might still find the codeframe of call outcomes too complex for direct implementation by their interviewers. Entering the codes into the grid of the contact form, might be too great an effort for the interviewers. Field organisations might therefore opt for only recording general outcomes in this grid on the contact form and route the interviewer to more detailed outcomes on additional pages. For example, the interviewer might code a general 'ineligible' in the grid and specify the reason for ineligibility on one of the next pages of the contact form. For final call outcomes and computerised contact forms this can be easily done. For non-final call outcomes on paper, however, care should be taken that the call attempt at which the detailed outcome was obtained is recorded and can be matched with the general outcome on the grid. Only if it is straightforward to obtain the complete call-level information from the routed contact form should this type of routing be applied. In the ESS for example there are a number of countries where the general information on the first page of the contact form can not be directly matched with more detailed breakdowns on later pages. For non-final call outcomes, especially, it is therefore preferable to directly implement all call-level outcome codes on in the contact form grid.

Nevertheless, survey organisations will differ as to which type of systems they use for collecting information from the field and how they manage their interviewers. They will therefore also have different practices as to which outcomes are collected directly in the grid and which are routed from general outcomes to more detailed outcomes. This might have to do with in-house survey traditions and which (and how many) of the call outcomes in Table 2 are applicable to a certain country or survey. One should be aware though that differences in the implementation of contact data might also affect the comparability of the results.

Table 2: Codeframe of call outcome codes
(for in-person surveys of one individual interviewed per household)

Call outcome	**Sampling frame** Only sample type AD = addresses HH = households IN = individuals	**Type of outcome**
Ineligible		
Address not eligible	HH/AD	final
Not yet built/under construction	HH/AD	final
Demolished/derelict	HH/AD	final
Vacant/empty	HH/AD	final
Non-residential address (business, hospital, school etc)	HH/AD	final
Address out of sample (i.e. not within country)	AD	final
Other	HH/AD	final
Address occupied, but no eligible household	HH/AD	final
Vacation homes	HH/AD	final
Temporary residences	HH/AD	final
Other	HH/AD	final
Household occupied, but no eligible person	HH/AD	final
Nobody of eligible age in household	HH/AD	final
Sample person not eligible (e.g. not aged 18, moved abroad)	IN	final
Other		final
Other ineligible		final
Non-contact		
No access to the address		non-final
Area not accessible		non-final
No contact made with anyone at the address, no access to household		non-final
Housing block with access code / Gated community		non-final
Refusal at the address, no access to household		non-final
No contact with the household		non-final
Contact at the household, but not with household member		non-final
Non-household member indicated good time for reaching household (soft appointment with household)		non-final
Non-household member indicated good time for reaching sample person (soft appointment with sample person)		non-final
Contact		
Contact with household, but		
Language problems		non-final
Too busy, not a good time to call		non-final
Household refusal		non-final*
Proxy refusal on behalf of sample person		non-final*
Sample person not available		non-final
Household member indicated good time for reaching sample person (soft appointment)		non-final
Appointment made with household member for sample person (hard appointment)		non-final
Appointment made with sample person (hard appointment)		non-final
Contact with the sample person, but		
Language problems		final*
Physically or mentally unable		final
Too busy, not a good time to call		non-final
Soft refusal to participate		non-final
Hard refusal to participate		final*
Interview		
Completed interview		final
Broken-off interview, to be completed		non-final
Refusal during interview, interview not to be completed		final

The first column of the codeframe in Table 2 describes the call outcome; the second column indicates again if the outcome is restricted to a certain type of sampling frame; the third column indicates whether a call outcome is final or non-final. For some codes the survey organisation might feel that a final outcome in this codeframe is actually non-final in their country or survey setting. These codes are indicated by an asterisk. For example, for surveys that are carried out in several languages, a sample unit having language problems might simply mean that an interviewer speaking the sample unit's language needs to be sent to do the interview. In this case 'contact with the sample unit, but language problems' would be a non-final call outcome.

From call outcome to case outcome: deriving a final case outcome

Those call outcome codes that are collected during fieldwork on a contact form and by means of the call-level codeframe need to be assigned to a case outcome code that matches the codes in the case-level codeframe. For final call outcomes this is straightforward, because the final call outcome in the contact form equals the final case outcome. The categories in the call-level and the case-level codeframes match to account for this. The non-final call outcomes, however, need to be coded into a derived final case outcome. Two issues need to be considered here: (1) the integration of information on the call outcomes with information on whether the household and the sample unit were selected (for samples of addresses and samples of households) and (2) the process of coding a sequence of several contact attempts into one final case outcome.

Before coding a sequence of several contact attempts into one final case outcome the survey manager or researcher has to decide on a coding strategy. The three most commonly used coding strategies are (1) using the last outcome of a sequence of calls (most-recent coding), (2) assigning the final case outcome according to a priority ranking of call outcomes (priority coding) and (3) subjective coding. With most-recent coding, the outcome of the last call to a sample unit is defined as the case outcome (AAPOR, 2000). Accordingly, the intermediate call outcomes in a sequence of contact attempts are irrelevant for this coding strategy. As discussed above some case management systems only save the last call outcome and thereby automatically prescribe most recent coding. Priority coding, on the other hand, involves arranging call outcomes according to a priority ranking, in which some outcome codes take

priority over others. The final case outcome of a sample unit equals the non-final call outcome with the highest priority (Lynn et al., 2001). Finally, subjective coding refers to situations where the rules for assigning a sequence of non-final call outcomes to a case outcome are not defined. Typically in such situations only descriptions of each case outcome code (which can vary in their precision) 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 (Blom, Lynn and Jäckle, 2008, p.10).

Survey organisations working with issue outcomes also have to take a decision regarding which coding procedure is used to derive an issue outcome. In addition, they need to determine how to assign final case outcomes if a case has more than one issue outcome. The same coding strategies are available for deriving a final case outcome from several issue outcomes.

The choice of coding strategy for deriving case outcomes may vary across survey organisations. When comparing response outcomes across interviewers, surveys or countries the same coding strategy needs to be chosen. Subjective coding is unsuitable, as this renders it impossible to know whether differences in response outcomes across surveys are a result of the differences in actual responses or are due to differential subjective coding. Instead researchers need to decide on whether to use most-recent or priority coding. If priority coding is chosen they will further need to agree on one priority order that is applied to all units of comparison. In the case of organisations working with issue outcomes care should be take that the call outcomes of all issues are taken into account when deriving final case outcomes. Section three describes the effects of most-recent and priority coding on the derived case outcomes at the example of the ESS.

Survey organisations may also work with simpler code frames and subjective outcome codes during fieldwork (especially when call outcomes are collected by paper-based technology), but examine a more elaborate codeframe when examining data quality after fieldwork has finished. Because keying and coding outcomes takes time, additional subjectively coded issue outcome codes might be necessary while

fieldwork is in progress. For comparability of response outcomes, however, standardised case outcomes should be derived once fieldwork is finished.

SECTION 3: THE EFFECT OF CODING STRATEGIES ON CASE OUTCOMES

The third section looks at the impact of the coding strategy on the comparability of fieldwork response outcomes using the contact data of the first round of the ESS. Having established that subjective coding does not allow for comparisons of case outcomes, only priority coding and most-recent coding are examined. The analysis starts with a theoretical account of the effects of the coding strategies with different types of call outcome sequences. I then continue with a description of the ESS contact data and compare the case-level outcome codes in the ESS to those in my codeframe in section 2. Finally, case outcomes are derived from the ESS call-level contact data by means of priority coding and by means of most-recent coding. Differences in case outcomes, when using the two coding strategies are examined and the implications for comparative response analyses discussed.

Comparing most-recent and priority coding conceptually

The rationale behind choosing priority or most-recent coding for deriving case outcomes works along two dimensions: (1) ease of implementation and (2) accuracy in reflecting the de facto response outcomes. Most-recent coding has clear advantages over priority coding in terms of ease of implementation. Deriving case outcomes according to a priority ranking is more complicated than simply assigning the outcome of the last contact attempt to the case outcome. Especially if the call record is kept on paper, survey agencies might be wary of implementing the priority coded derivation, though interviewers could, of course, be instructed to follow priority rules. As discussed in section 1, some survey organisations work with computerised systems that only record the outcome of the last contact attempt. This means that the case outcome can only be assigned via most-recent coding. On the other hand, computerisation also means that a simple algorithm can be built into the CAPI script. This then automatically derives the case outcome from the call-outcome sequence by means of a priority ranking. Therefore, most-recent coding is easier to implement than

priority coding, however, with interviewer instructions or an algorithm in a computerised contact form, deriving a priority coded case outcome is definitely possible.

However, how do we reflect the de facto response outcomes most accurately? I take the position that the final case outcome should be an indicator of the extent to which the aim of the survey process was achieved. In the survey process an interview with the sample unit is the ultimate aim. Therefore, the more accurately a case outcome reflects the extent to which the survey process was close to achieving an interview, the more accurate an indicator of the de facto response outcome it is. As I will show in the following, a sensibly ranked priority coding can take this into account. Depending on the sequence of contact attempts, however, most-recent coding does not accurately reflect how close to achieving an interview the interviewer was. Consequently, to reflect de facto response outcomes as accurately as possible, priority coding is the method of choice; at least if a sensible priority order is chosen.

I now introduce what such a priority order might look like. Usually only sequences of non-final call outcomes need to be coded into case outcomes. Therefore the general call outcomes of relevance here are uncertain eligibility, non-contact, refusal by sample unit, proxy refusal, appointment and other contact.[2] Since uncertain eligibility is a state in the fieldwork process that needs to be further defined, any other call outcomes should take priority over this outcome. A non-contact case outcome usually means that this sample unit has never been contacted at all and the interviewer has not had any chance to voice her request. Therefore, given the eligibility of the sample unit was established, all other outcomes should take priority over non-contact. A proxy refusal indicates that it was possible to voice the request for an interview to a responsible household member. A refusal by the sample member, however, assumes that the request has been made to the sample unit directly. And an appointment indicates that achieving an interview was within reach. Proxy refusals, refusals by the sample unit and appointments are each steps towards achieving an interview. A refusal by the sample unit should therefore take priority over a proxy refusal and an

[2] For reasons of simplicity only general call outcome categories are listed here, though all categories (down to level-3) have to be considered when designing a priority ranking. In the details of the priority order researchers might have different convictions of which outcome they deem more important, i.e. which category receives a low/high priority.

appointment for interview should take priority over any kind of refusal.[3] Whether or not other contact call outcomes (e.g. mentally/physically unable or language problems) should have priority over the refusal outcomes is a difficult issue. However, usually one considers refusals (and appointments) to have priority over other contact outcomes. The resulting priority ranking of these general call outcomes then is as follows.

(broken) appointment
⇩
refusal by sample unit
⇩
proxy refusal
⇩
other contact without interview
⇩
non-contact
⇩
uncertain eligibility[4]

How does the case outcome derived by this type of priority ranking differ from one derived by means of most-recent coding? For illustration I consider different kinds of scenarios with sequences of three call outcomes.

Sequence 1: refusal conversion attempts

Contact attempt number	1	2	3
Call outcome	non-contact	refusal	non-contact

Case outcome (priority): refusal
Case outcome (most-recent): non-contact

In case of refusal conversion attempts, where the refusal is not converted into an interview and never reached again, priority coding will derive a refusal case outcome while the most-recent coding derives a non-contact. The priority case outcome therefore reflects the fact that at some point this sample unit was successfully contacted, i.e. that the contacting procedure has been fruitful, yet the procedure for gaining cooperation has not. The non-contact case outcome derived by means of most-recent coding does not reflect these issues.

[3] The researcher should further be aware that an appointment that is not turned into an interview by the end of fieldwork, is usually considered a broken appointment and therefore a refusal.
[4] ⇩ means 'has priority over'

Sequence 2: broken appointment

Contact attempt number	1	2	3
Call outcome	appointment	non-contact	non-contact

Case outcome (priority): broken appointment / refusal
Case outcome (most-recent): non-contact

Similarly, if ever any appointments are missed or broken, priority coding gives a better picture of the survey process. A sensible priority ranking will identify unsuccessful attempts of turning an appointment into interviews as broken appointments, i.e. refusals, while most-recent coding might identify them differently (in sequence 2 as non-contact).

In summary, the main disadvantage of most-recent coding is that it identifies a case as a non-contact, if the last contact attempt was a non-contact, even if contact was achieved at an intermediate contact attempt. It therefore fails to reflect the extent to which an interviewer was close to achieving an interview.

Sequence 3: drop-off questionnaires

Contact attempt number	1	2	3
Call outcome	appointment	interview	non-contact

Case outcome (priority): interview
Case outcome (most-recent): non-contact

Sequence 3 displays a special case. Here a contact attempt was made, even though an interview was already achieved. The reason for this might be that the interviewer still has to complete a few questions, yet does not manage to make contact again. Or the survey contains drop-off questionnaires, which the interviewer leaves with the sample unit after the interview and which she later attempts to collect. In this sequence, the collection of the drop-off questionnaire would have been unsuccessful. As a consequence the case outcome derived by means of the most-recent coding is a non-contact, even though an interview was achieved. Usually, these kinds of situations do not pose any serious problem, since the survey organisation holds a record of the achieved interviews and will derive case outcomes and calculate outcome rates accordingly. Though it can be a problem to the methodologist if full call data are not made available, e.g. if you only have the summary variables ‘total number of calls’ (3) and ‘case outcome’ (interview).

Sequence 4: other outcomes

Contact attempt number	1	2	3
Call outcome	non-contact	proxy refusal	language problem

Case outcome (priority): proxy refusal
Case outcome (most-recent): language problem

Sequence 4 is an example of a more difficult case and displays that some degree of arbitrariness is involved in any kind of case outcome coding, including any priority ranking. The call outcome sequence 3 can be the result of two different possible circumstances. On the one hand the sequence might represent a situation, where the interviewer is refused at the door by a household member. At a later contact attempt the interviewer finds out that the reason for the refusal was that the sample unit does not speak the language of the interview and was thus reluctant to speak to the interviewer. The case outcome of this sequence of call attempts should thus be 'language problem'. The sequence might on the other hand depict a situation in which at the later contact attempt the sample unit only claims not to speak the language to avoid being contacted again. Therefore the correct case outcome should be a refusal. The sequence therefore illustrates that, to some extent, the choices made in the priority ranking are necessarily arbitrary, because the same sequence of call outcomes can have different causes. Unless additional information on the nonresponse process is available, this problem cannot be solved. It is assumed, however, that in each survey only a small number of cases are affected by such problems, such that it is negligible.

In summary, this account of the possible effects of the various coding strategies has shown that priority coding will yield a more accurate picture of the de facto response outcomes. Since most-recent coding focuses on the last contact attempt only, it will over-report cases that are at the bottom of the priority sequence, especially non-contacts. Nonetheless, the researcher should keep in mind that also with priority coding there is always some degree of arbitrariness. Having established the effects of priority and most-recent coding in theory, I now turn to the implications in an empirical analysis of the ESS response outcomes.

The ESS contact data and outcome codes

In 2002 the ESS became the first cross-national survey to collect and make publicly available standardised call-level contact data. Since the ESS is a cross-national survey with national fieldwork implementation, fieldwork control is primarily carried out at the national level. Collecting these cross-national contact data therefore had several aims, including "to monitor and supervise the fieldwork process [at the cross-national level], to check whether the fieldwork had been carried out according to the [ESS] specifications, to compute nonresponse rates, contact rates and cooperation rates, and to estimate the bias due to nonresponse." (Stoop et al., 2003, p.1) The ESS is a biennial cross-sectional survey and since 2002 the contact forms for collecting these contact data have been improved and updated before each round. Before the first round, ESS researchers studied contact forms from eleven organisations in six different countries. "In the end, these efforts resulted in a standardised contact form specification and the construction of a standardised data file" (Billiet et al., 2007, p.140) . The ESS contact forms collect both call- and case-level information on response outcomes, nonresponse processes and neighbourhood characteristics.

Table 3: Information collected in the ESS contact forms

Call-level	• Response outcome of each contact attempt • Information on selection procedure (for household and address samples) • Day, date, month and hour of each contact attempt • Mode of visit • Reason for refusal • Interviewer identification
Case-level	• Neighbourhood characteristics of each sample unit • Estimated age and gender of sample person at refusal

Section two describes the collection and derivation of response outcomes as a top-down approach. This means that the codeframe of final case outcomes is defined first; the section then lays out ways of arriving at these final case outcomes by means of in-office codes and final and non-final call outcomes implemented in a contact form in the field. In the ESS the collection of response outcomes is a bottom-up approach. The ESS centrally develops model contact forms. These model contact forms can be directly implemented in the field.[5] The content of the model contact forms together with the contact data file specifications reflects which variables interviewers in each

[5] Countries were allowed to use different contact forms, provided that they collected all the required call- and case-level response and auxiliary data. The model contact forms, nationally implemented contact forms and the contact data are available from the 'ESS Data' tab on the ESS website (www.europeansocialsurvey.org).

country are expected to collect. As a consequence, the ESS does not explicitly distinguish between in-office, final and non-final outcomes, though by the very nature of these codes many countries implicitly make this distinction. This means that the ESS prescribes the call-level outcomes that interviewers need to collect; the final case outcomes are then derived from these call outcomes. On the first page of the ESS contact form, the interviewer registers the general outcome code of a call. The general outcome codes in rounds 1 to 3 of the ESS are listed in Table 4.

Table 4: General (non-routed) outcome codes in the ESS

Outcome Code	Round 1	Round 2 / 3
Interview (partial or complete)	X	
Completed interview		X
Partial interview		X
Contact with respondent but no interview*	X	
Only contact with someone else*	X	
Contact with someone, target respondent not yet selected*		X
Contact with target respondent but no interview*		X
Contact with somebody other than target respondent*		X
No contact at all	X	X
Address not valid (unoccupied, demolished, not residential…)	X	X

* For samples of individuals these general categories include routing to ineligible outcomes for individuals (e.g. deceased, moved out of country)

For each contact attempt the interviewer is then routed to the next pages of the contact form where more detail is collected about these general outcomes, for final call outcomes (i.e. ineligibles) as well as for the non-final call outcomes. Table 5 lists all call outcome codes that were required to be collected in each country and made available in the ESS contact data files of rounds 1 to 3. Since countries worked with different sampling frames, the table indicates which codes apply to which sampling strategy.

As mentioned earlier, routing non-final call outcomes is tricky as it means that the interviewer must carefully record the detailed call outcome, such that the general and specific outcomes can be matched. The ESS contact form allows for this, however, unfortunately, not all interviewers in the various countries filled in the contact forms accurately. In a significant number of cases interviewers failed to specify the general outcome of a contact attempt or specified it at the wrong call number, which means that the general and specific outcomes cannot be matched. As a consequence, during the matching an additional category “undefined” as subheading under the main categories of interview, contact, non-contact and ineligible had to be created. These

categories capture contact attempts where the general call outcome could not be further specified with a specific call outcome. Therefore, in addition to the call-level codes that Table 5 lists, the additional codes ‘interview – undefined’, ‘contact – undefined’, ‘non-contact – undefined’ and ‘ineligible – undefined’ need to be considered.

Table 5: Routed outcome codes in the ESS

Outcome Code	Round 1	Round 2 / 3
Ineligible / address not valid		
Derelict or demolished house / address	All	All
Not yet built / not yet ready for occupation	All	All
Address is not traceable, address was not sufficient	All	All
Address is not residential: only business / industrial purpose	All	All
Address is not residential: Institution	All	All
Address is not occupied (empty, second home, seasonal)	All	All
Respondent deceased	HH, IN	All
Respondent moved out of country	HH, IN	
Respondent moved out of country or to unknown destination		HH, IN
Other	All	All
No contact at all		
Nobody at home	All	
Broken appointment	All	
At home but did not answer the door	All	
Could not obtain access to housing unit	All	
Respondent / household moved	All	
Other	All	
No contact at all		All
Contact, but no interview; only contact with someone else		
Appointment	All	All
Refusal of respondent	All	All
Refusal by someone else (by proxy), on behalf of the respondent	All	All
Household refusal (before selection)		HH, AD
Refusal. Don’t know if target respondent		IN
Respondent Is unavailable / not at home until ….		All
Respondent is temporarily unavailable but will be available before end of the fieldwork period (e.g. out, away, on holiday, sick)	All	
Respondent is unavailable throughout the fieldwork period for other reasons (e.g. away, abroad …)	All	
Respondent is mentally or physically unable to co-operate throughout the fieldwork period	All	All
Respondent moved, still in country	HH, IN	HH, IN
Language-barrier of respondent	All	All
Other	All	All
Interview		
Interview completed	All	All
Interview broken off or incomplete, to be completed at a later date	All	
Interview broken off or incomplete, will not be completed	All	
Partial interview		All

Overall the ESS contact forms and contact data file allow the recording of up to ten contact attempts made to a sample unit. However, some of the specific outcome codes allow up to ten specific outcomes (i.e. contact outcomes (including refusals) and non-contact outcomes in round 1), while other general outcomes can only be followed by specific outcomes up to three times (i.e. interviews in round 1) or once only (i.e. ineligibles). Finally there are general outcomes that are not further specified at all (i.e. non-contact outcomes in rounds 2 and 3 and interview outcomes in rounds 2 and 3). Indirectly this also implies that ineligibles were recorded as final call outcomes, since only one specific ineligible call outcome could be recorded. In contrast to my codeframe, the ESS treated interviews as non-final call outcomes in round 1, since it was assumed that an interview could be started at one call and continued at another call. The contact forms recorded these as separate contact attempts, rather than a continuation of the interview.

Outcomes that are solely received in the office do not fit the standardised ESS contact form or data file specification. However, also in ESS fieldwork there will be instances when a case has an in-office outcome only. In round 3 of the ESS, for example, Cyprus ended fieldwork prematurely and 381 sample units were never issued to the field. Consequently, these cases were also missing from the contact data. They were added to the contact data set afterwards, however, as 'empty' cases only, i.e. any information on the field process and neighbourhood was missing. Similarly in Denmark, which uses the population register as a sampling frame, a new data protection procedure was put into place in 2006, making it possible for people to block their registry entry from being contacted for research purposes (including academic and government research) in form of a universal opt-out procedure. In the third round, 360 selected sample units had opted out in this way and could therefore not be issued to an interviewer. Again these cases were added to the contact data as 'empty' cases. However, since the ESS field outcomes codeframe is a call-level codeframe it does not allow for any in-office case outcomes and no separate outcome codes are available. Instead, the ESS team treated these cases as non-contacts, since no contact attempt was ever made or achieved.[6]

[6] Other options for treating these cases could have been: (1) office refusal, since the sample units refused participation, when registering, (2) ineligible, since they might be treated as not part of the

The ESS also collects information on whether a contact attempt resulted in an appointment. Furthermore, the general outcome codes (Table 4) provide information regarding with whom contact was made. From this one can derive whether an appointment was made with the sample person or with someone else. The ESS contact forms assume that the interviewer knows whom they are talking to at the door (or via the intercom). However, oftentimes interviewers do not, which means that this information is unfortunately not very reliable. In addition, it is not clear whether the appointment made is a general indication of when a sample person might be available, i.e. a soft appointment or whether it is a hard appointment for an interview at a fixed date and time. Consequently, compared to the codeframe presented in section two the information on appointments in the ESS is quite general.

Regarding cases of uncertain eligibility the ESS model contact forms do not allow interviewers to record when they are unsure about the eligibility of a case. This can be problematic, especially in countries where the sampling frame is erroneous or where researchers have to rely on sampling addresses via random-route procedures. Since the eligibility of the address, the household and the individual has to be established during the contacting process, there will always be cases of uncertain eligibility. As mentioned earlier, because the ESS samples individuals aged 15 and over, one can assume that each eligible household contains at least eligible sample unit. However, the interviewer might be unable to establish whether the address contains a resident household. For example, they might be uncertain if a house is a holiday apartment only (i.e. ineligible) or occupied throughout the year (i.e. eligible). One of the invalid outcome codes in the ESS, address not traceable, is especially problematic. In theory, addresses that are not traceable exist, but the interviewer cannot locate it. However, a non-traceable address might also not exist (not yet or not anymore), for example if the sample is drawn from an inaccurate postal list.

For samples of individuals an interviewer might also be unable to ascertain whether the sampled person is resident in the country. If someone moves abroad long-term they are not eligible to be interviewed. Especially in Eastern European countries the

target population, (3) uncertain eligibility (never issued) or (4) a separate, country-specific response outcome.

enlargement of the EU means that a considerable proportion of the people on the population registers (and therefore the sampling frame) live and work in Western Europe several months a year or have moved abroad completely. If an interviewer is unable to make contact with anyone at the address or is refused this information, they will not be able to establish if the sampled person actually still resides in the country and is eligible or not.

In summary, the ESS codeframe of field outcomes differs from the codeframe that I propose in section two with regards to some important aspects. Most notably all outcomes are collected by means of call-level outcomes and the ESS does not provide codes for cases of uncertain eligibility or in-office outcomes. The information collected regarding appointments is less comprehensive in the ESS contact forms. Finally the collection of call outcomes is error-prone, since it is routed from general outcomes to more specific ones that need to be matched. Consequently, not all of the codes in my case-level codeframe can be derived from the ESS call outcomes. Nonetheless, the ESS is a pioneer in collecting comparative fieldwork outcomes and is an important source of cross-national data on response outcomes and processes. Due to the uniformity of the contact forms, the collected response outcomes and other auxiliary data, the ESS contact form data are, though not ideal, well-suited to comparative analyses of cross-national survey nonresponse.

Deriving case outcomes for the ESS

At the beginning of this section I described the theoretical impact of most-recent and priority coding on case outcomes when different kinds of call outcome sequences are considered. In the following I examine the differences ESS case outcomes when comparing between case outcomes derived by means of priority coding with those derived by mean of most-recent coding. To set the scene, I first regard the case outcomes that can be derived from the ESS contact data in round 1. I describe errors found in the ESS contact data and why errors differ between coding strategies. Subsequently, I investigate differences in case outcomes for a pool of all countries. This is followed about a more detailed analysis of differences across ESS countries in refusal, non-contact and ineligible case outcomes. Finally, the analysis investigates the impact on the comparability of outcome rates across ESS countries of not specifying the coding strategy to be used for deriving case outcomes.

In rounds 1 and 2 the central coordination of the ESS did not instruct the national teams in how to derive case outcomes from the call outcomes in the contact forms. As a consequence, countries tended to use different, undocumented techniques for deriving final case outcomes. Whilst the outcome of each case of the gross sample is not documented, the national coordinators of the ESS did report distributions of relevant case outcomes in a standardised technical report (ESS, 2003). Since round 3 the ESS team provides countries with an algorithm for deriving case outcomes from ESS call outcomes.[7] As a consequence, the comparability of case outcome distributions across countries should have improved since round 3.

In the following I use the round 1 ESS contact data to compare the differential impact of a strict priority-ranked coding and most-recent coding for deriving case outcomes on the distribution of these outcomes in 16 countries participating in round 1 of the ESS[8]. With the most-recent coding system the outcome of the last contact attempt determines the final case outcome. For a derivation of case outcomes by means of priority coding the call outcomes were arranged according to the hierarchy in Table 6. For each sample unit the call outcome with the highest ranking on the hierarchy determines the final case outcome. In the process of deriving case outcomes from call outcomes, the label of an outcome code might need to be changed. For example a call-level outcome 'appointment' that is not followed up by an 'interview', becomes a 'broken appointment' (i.e. an implicit refusal) in the final case outcome. The first column of Table 6 ranks the priority of call outcomes, with the top category

[7] This algorithm is based on the case outcome coding of Billiet et al. (2007), who use a combination of priority and the most-recent coding. "[T]he outcome of the last contact was used as the final nonresponse code. The exception was when a refusal occurred at an earlier visit and subsequent contacts with the household resulted in other eligible nonresponse outcomes. In this case, the final nonresponse code was "refusal to participate""(Billiet et al., 2007, p. 142). Since the case outcomes derived by Billiet et al. (2007) have a different purpose, they only derive case outcomes at approximately level-1 detail. A direct comparison of these case outcomes with those derived further below in this paper is not possible.

[8] See Table 8 for countries that were included in the analysis and sample sizes. Six countries that participated in round 1 of the ESS are not included in the analysis for the following reasons:
Czech Republic: Incomplete contact data. Contact data of 619 sample units not issued.
Denmark: Incomplete contact data. The first contact attempt is missing.
France: No contact data submitted in round 1
Israel: Inconsistencies between data specification and actual data. Contact data and main interview data are partially incompatible.
Sweden: No contact data submitted in round 1
Switzerland: Response experiment with extensive telephone contacting. Very high numbers of recorded contact attempts make the Swiss data incompatible with the ESS data format.

displaying the outcome with the highest priority and the bottom category that with the lowest priority. Column two lists the respective case outcomes. Note that the same change in labelling takes place when deriving case outcomes by means of most-recent coding. Furthermore, the first part of each column indicates which level-1 outcome code the outcome belongs to, i.e. whether it is an interview, ineligible, a refusal, other contact or non-contact.

Table 6: Priority ranking of response outcomes

Call outcomes		**Case outcomes**	
interview	completed	interview	completed
interview	broken off / incomplete tbc at later date	refusal	interview broken off / never completed
interview	broken off / incomplete not tbc	refusal	interview broken off / never completed
interview	undefined	refusal	interview undefined, no record in main data
ineligible	not residential / institution	ineligible	not residential / institution
ineligible	not residential / business	ineligible	not residential / business
ineligible	not yet built	ineligible	not yet built
ineligible	derelict	ineligible	derelict
ineligible	not occupied	ineligible	not occupied
ineligible	not traceable	ineligible	not traceable
ineligible	other	ineligible	other
ineligible	undefined	ineligible	undefined
contact	respondent deceased	ineligible	respondent deceased
contact	respondent moved abroad	ineligible	respondent moved abroad
contact	refusal by respondent	refusal	refusal by respondent
contact	refusal by someone else	refusal	refusal by someone else
contact	appointment made with respondent	refusal	appointment w respondent, never realised
contact	appointment made with someone else	contact	appointment w someone else, never realised
non-contact	broken appointment	refusal	broken appointment
contact	mentally / physically unable	contact	mentally / physically unable
contact	language barrier	contact	language barrier
contact	unavailable throughout fieldwork period	contact	unavailable throughout fieldwork period
contact	temporarily unavailable	contact	temporarily unavailable
contact	other eligible contact no interview	contact	other eligible contact no interview
contact	undefined	contact	undefined
contact	respondent moved, still in country	non-contact	respondent moved, still in country
non-contact	respondent/household moved, new address	non-contact	respondent/ household moved new address
non-contact	at home but no answer	non-contact	at home but no answer
non-contact	nobody at home	non-contact	nobody at home
non-contact	no access to housing unit	non-contact	no access to housing unit
non-contact	other	non-contact	other
non-contact	undefined	non-contact	undefined
missing	call outcome missing	missing	case outcome / all call outcomes missing

Note that ineligibles and interviews, which both have final call outcomes that can directly determine the case outcome, are nevertheless included in the hierarchy and derived from the call-level outcomes. The reason for this is that the ESS contact forms do not explicitly distinguish between final and non-final call outcomes and in-office

case outcomes. As a consequence, interviewers were able to register another contact attempt after a final call outcome (e.g. interview or ineligible).
Like any type of survey data, contact data are prone to errors. Lacking edit checks, interviewer training on filling in contact forms and simply sufficient attention by survey researchers, some contact data available for analysis contain errors in the outcome codes. This is especially the case in cross-national surveys where contact data are collected in countries that have little experience with this type of data or when contact data are only collected solely as a by-product of survey operations (without considering the reliable deduction of outcome rates or methodological research).

With respect to response outcomes, an important source of error is that interviewers do not finish completing their contact form once an interview is achieved. From a survey operations point of view this makes sense, since once an interview is transmitted, no further action needs to be taken on the case (except for back-checking). Yet this also means that when deriving final case outcomes the researcher will find fewer interviews in the contact data than there are cases in the main interview data file.

Because sample units in each country in the ESS contact data and the ESS main interview data have the same unique identifier, it is possible to pinpoint which sample units in the contact data have an interview in the main data file. I use this information to correct a non-interview code in the final case outcome into a completed interview. The ESS does not distinguish between partial and complete interviews in the main data file, therefore all sample units with an interview in the main data file are considered to have completed the interview. Similarly cases that have an 'interview completed' final outcome code in the contact data but no interview in the main data file can be thus recoded into 'interview undefined'.

Out of the 16 countries included in this analysis 12 countries had some cases where the final case outcome would not have been identified as an interview had the researcher relied on the contact data only (Tables A1 and A2, Appendix). In total these are 449 misclassified cases with priority coding and 532 misclassified cases with most-recent coding. The fact that there are more misclassifications associated

with most-recent coding implies that there are cases where a call outcome of interview is followed by extra call(s) with a different outcome. This seems to be largely due to interviewers re-contacting a sample unit to collect the drop-off questionnaire. Two countries show a much larger number of misclassifications with most-recent than with priority coding: Finland and the UK. These are both countries that used drop-off supplementary questionnaires which interviewers had to collect later.[9]

Across coding strategies, the most frequent reason for misclassifications of interviews was that the general interview code in the contact data was not further specified. Consequently no completed interview was recorded. Yet also many cases with broken-off interviews, appointments, refusals and missing case outcomes in the contact data had an interview in the main data. There are large differences across countries in the number of misclassifications of interviews in the contact data. Apparently, in some countries interviewers recorded the outcome of the call leading to an interview more accurately than in others.

Though less frequently there were also cases, where the contact data had recorded a completed interview, yet no such data exist on the main interview file (14 cases priority coded and 12 cases most-recent coded).

Priority and most-recent coded outcomes in the ESS

In the following I present ESS round 1 final case outcomes derived by means of priority and most-recent coding. For reasons of clarity the outcomes are first presented at level-1 detail across all 16 countries. Differences across countries are analysed thereafter. For reasons of complexity tables of level-2 details are not displayed. However, findings from analyses of lower-level outcomes are discussed in the text where appropriate.

[9] The ESS questionnaire contains a so-called supplementary questionnaire, which can be either implemented as part of the main face-to-face interview or as a self-completion questionnaire. If a country chooses to implement the supplementary questionnaire as a self-completion questionnaire, the interviewers can either (a) wait for the respondent to fill in the questionnaire, (b) let the respondent fill in the questionnaire in their absence and mail it back or (c) leave the questionnaire with the respondent and collect these drop-off questionnaires again in person.

The rows of Table 7 display the frequency of case outcomes derived by means of the last call outcome (i.e. through most-recent coding); the columns contain the outcomes derived via the priority ranking of call outcomes described in Table 6 (i.e. through priority coding). The diagonal of the table shows the number of cases where most-recent and priority coding lead to the same outcome at level 1. The analysis is aggregated over the 16 countries. The sample size for each country is displayed in Table 8.

Table 7: Level-1 case outcomes: most-recent versus priority coding - round 1

final outcome (most-recent coding)	**final outcome (priority coding)**						
	Interview	Ineligible	Refusal	Other contact	Non-contact	Missing	Total
Interview	31,447	0	0	0	0	0	31,447
Ineligible	0	2,083	0	0	0	0	2,083
Refusal	0	35	12,332	0	0	0	12,367
Other contact	0	11	407	2	0	0	2,771
Non-contact	0	27	866	403	3,382	0	4,678
Missing	0	0	8	10	4	106	128
Total	31,447	2,156	13,613	2,766	3,386	106	53,474

For both coding strategies I assigned a ‘completed interview’ outcome, if the main ESS data file contained an interview, but the contact data did not. Similarly I assigned an ‘interview undefined’[10] in cases where the interviewer had coded a ‘completed interview’ in the contact data, but no interview was available for that case on the main data file.

With priority coding there are considerably fewer missing case outcomes than with most-recent coding (106 compared to 128). The reason for this is that priority ranking only codes a case as ‘missing’ if all call outcomes are missing. Most-recent coding in contrast also derives a missing case outcome if only the last call outcome is missing, though previous outcomes might not be missing. Therefore, we lose information about the whole case, if the last call outcome is missing. This of course happens only if some contact data are recorded for the last call, but not the outcome. If the last call is not recorded at all, then the previous call will appear as the last call. This loss of information with most-recent coding could be prevented, if this coding strategy was adjusted to code the last non-missing call outcome as the case outcome instead of just the last call.

[10] The level-2 outcome ‘interview undefined’ is classified as a refusal at level 1.

The most significant differences between the two coding strategies can be found in the refusal, non-contact and other contact outcomes. Regarding refusals priority coding registered 10.1% more refusals case outcomes than most-recent coding (13,613 cases priority coded compared with 12,367 cases most-recent coded). This is due to the fact that any unsuccessful attempt at converting refusals will be registered as a refusal with priority coding, while if the last call outcome is taken, the case outcome might also be another code (mainly non-contact or other contact). Looking at the detailed outcome codes one further finds that most of these cases that were coded a refusal with priority coding were coded 'non-contact, nobody at home' with most-recent coding (670 cases, i.e. 52.3%). Interesting are also cases that were coded refusals with priority coding, but 'other contact' with most-recent coding. For 158 of these cases (12.3%) the 'other contact' outcome was 'unavailable temporarily' or 'unavailable during fieldwork period'. These are possibly cases, where a household member claimed that the target person was unavailable to avoid an interview. If this were true, the outcome of these calls would be disguised refusals. Another 119 cases (9.3%) that were coded refusals with priority coding were mentally or physically unable to do the interview at the last contact attempt. Again, the question is, whether these cases were actually unable to do the interview at the time of the call or whether this was just an easy way out.

With respect to non-contacts, on the other hand, most-recent coding derived 38.2% more non-contact case outcomes than priority coding (4,678 most-recent coded compared with 3,386 priority coded). Again this can be attributed to repeated, yet unsuccessful call-backs. In the priority ranking at hand a non-contact takes very low priority. If in a sequence of calls any contact was established with the household at any point and if this is then followed by non-contacts in subsequent calls, priority coding will assign a contact outcome to the case, while most-recent coding will assign a non-contact. In fact, most of these cases are refusal conversion attempts; of the cases that were coded non-contact with most-recent coding 378 (29.2%) were a refusal by the sample person when derived by means of priority coding, 298 (23.0%) were hard or soft broken appointments, and another 188 (14.5%) were refusals by someone other than the target person.

The proportion of cases coded ‘other contact no interview’ is approximately the same in both coding strategies, though the specific cases assigned to this code differ across the two coding strategies.

Interestingly also the ineligible codes differ between the two coding strategies. 73 cases that receive an ineligible outcome with priority coding receive an eligible outcome when most-recent coding is applied. Apparently, interviewers went back to cases that they had previously recorded as ineligible. This should not be possible. Either a case is ineligible, which means that no more calls should be made to the sample unit, or a case is eligible and the interviewer initially assigned the wrong code. In the latter instance, the call outcome should have been corrected by the interviewer or the operations team. Of these 73 cases, the majority occurred in just two countries. In Ireland there were 31 and in Germany 14 cases that were coded ineligible with priority coding but eligible with most-recent coding. The fact that priority and most-recent outcome coding can lead to a different number of ineligibles demonstrates two problems with the ESS contact forms. First, since no clear distinction between final and non-final call-outcome codes is made, interviewers can record another call outcome after they have recorded what should be a final call outcome. With more rigorous fieldwork control countries should be able to prevent this from occurring in the future. Second, since no uncertain eligibility codes exist in the ESS, interviewers have to decide whether a case is ineligible or eligible at the contact attempt. Including such outcome codes in the ESS contact forms might improve the number of followed-up ineligible outcomes.

Differences across countries

In a next step I look at how the impact of different coding strategies differs across countries in round 1 of the ESS. Three outcomes I investigated more closely: ineligibles (because they have an impact on the calculated response rate), refusals and non-contacts (because of the large differences in the distribution of these outcomes depending on coding strategy). Table 8 provides one column for each of these case outcomes. In each column the number of cases of this outcome (i.e. ineligibles, refusals and non-contacts) derived by priority coding and most-recent coding is displayed.

Table 8: Coding differences across countries (priority – most-recent coding)

	Country	Ineligible		Refusal		Non-contact		Base
		Priority	Most-recent	Priority	Most-recent	Priority	Most-recent	
Samples of individuals	Belgium	139	134	800	722	204	294	3340
	Germany	215	201	1548	1483	677	806	5796
	Finland	39	35	607	471	49	160	2766
	Hungary	75	71	366	336	193	242	2484
	Norway*	84	84	780	777	122	122	3215
	Poland	57	55	573	520	63	113	2978
	Slovenia	56	55	329	271	125	177	2175
Samples of households or addresses	Austria	99	99	1004	961	380	416	3828
	Spain	408	407	1142	1055	339	416	3702
	UK	289	285	1134	799	186	498	4013
	Greece	1	1	544	517	66	87	3227
	Ireland	86	55	720	696	296	360	3241
	Italy	160	160	1283	1236	154	221	3000
	Luxembourg	192	185	1292	1242	360	397	3773
	Netherlands	86	86	914	734	94	259	3570
	Portugal	170	170	577	547	78	110	2366

Notes:
* Norway did not provide detailed outcomes of the interview, non-contact and ineligible codes. The smaller differences between the coding strategies might be due to this.

In most countries, the difference in final ineligibles between priority and most-recent coding is small. The exception is Ireland, where 86 cases that were coded ineligible with priority coding, but 55 cases with most-recent coding.

The differential impact of the coding strategies is significantly more pronounced – across almost all countries – for refusals and non-contacts. For refusals the countries with proportionately the largest differences between priority and most-recent coding are the UK (1134 priority coded versus 799 most-recent coded cases), Finland (607 priority coded versus 471 most-recent coded cases), the Netherlands (914 priority coded versus 734 most-recent coded cases) and Slovenia (329 priority coded versus 271 most-recent coded cases). Regarding the number of differentially coded cases, non-contacts are similarly affected as refusals. However, since the non-contact rate tends to be much lower than the rate of refusals in the ESS, the proportional impact of the coding strategy is much larger with non-contacts than with refusals. The countries most strongly affected by a difference between priority and most-recent coding are Finland (49 priority coded versus 160 most-recent coded cases), the Netherlands (94 priority coded versus 259 most-recent coded cases), the UK (186 priority coded versus 498 most-recent coded cases) and Poland (63 priority coded versus 113 most-recent coded cases). Three of the countries are strongly affected by the impact of

differential coding strategies for both refusals and non-contacts. In Finland, the Netherlands and the UK apparently a large number of refusals were re-approached, but not contacted again by the end of fieldwork.

Surprisingly, the differential impact of coding strategies across countries does not seem to be related to the sampling frame that the countries use. One might have expected a difference due to an expected difference in the distribution of call outcomes associated with differences in sampling frames and fieldwork.

To summarise, final case outcomes derived by means of priority coding do differ from those derived by means of most-recent coding. Differences between the coding strategies with regards to ineligible outcomes are due to ineligible codes having been recorded as non-final call outcomes which were followed up with another call attempt leading to an eligible outcome. This should be easily overcome by better field control, data editing and by including outcome codes for registering uncertain eligibility. Independent of such actions, however, large differences between the two coding strategies in the number of cases that were coded as refusal, other contact and non-contact occurred in the ESS contact data. Case outcomes derived by means of priority coding were more likely to be refusals but less likely to be non-contacts and the cases coded as other contact no interview differed across coding strategies. As described at the beginning of this section, these differences are likely to occur when elaborate refusal conversion attempts are undertaken, of which some were unsuccessful, as was the case in many countries in the ESS.

The differential impact of nonresponse due to refusal and nonresponse due to non-contact has received considerable attention in recent years (Lynn et al., 2002, Lynn and Clarke, 2002). In light of the findings here it is quite surprising that consistent derivation of case outcomes from call outcomes has not found more attention in the research community. If we believe that a non-contacted case creates different nonresponse bias from a refusal we should define more clearly how this outcome was achieved. The country comparisons have further shown that the impact of differential coding differs across countries. While in some countries the differences in coding have only a marginal impact, in other countries (especially Finland, the Netherlands

and the UK) the proportion of refusals and non-contacts is strongly affected by the coding strategy chosen.

The impact of differential coding on outcome rates

To round off the analyses, I look at whether differential coding has an impact on some of the typical outcome rates provided by survey organisations: the response rate, the contact rate and the cooperation rate. Since the ESS uses the same call outcome codes across countries and since for each country I have coded these call outcomes with the same coding strategies into final case outcomes, I can now estimate comparative outcome rates.

Unfortunately, estimating outcome rates is not that straightforward. Groves et al. (2004) mention three complications, two of which are also applicable here. First, in surveys where there are cases of uncertain eligibility it is not straightforward how to calculate outcome rates. As discussed earlier, registering cases of uncertain eligibility and providing response rate bands would be a solution for this. However, ESS does not specify any outcome codes for cases with unknown eligibility; interviewers and survey organisations have to decide whether a case was eligible or not. Therefore, uncertain eligibility cannot be taken into account in my calculations. Second, "unequal probabilities of selection are assigned to different elements in the sampling frame (e.g., oversamples of minority ethnic groups). In this case, it is unclear whether selection weights should be used in the computation of response rates" (Groves et al., 2004, p. 183). Whether selection weights are considered should depend on the purpose of the outcome rate calculated (cf. Lynn et al., 2002). If the outcome rates are to provide insight in fieldwork efficiency, no weighting for probability of selection needs to be carried out. If, however, the outcome rates are to tell us something about possible nonresponse bias, weighting for the probability of selection will be necessary. In round 1, the ESS countries only provided information on selection probabilities for responding sample units. Therefore, it is impossible to weight the outcome rates for the probability of selection of the sample units.

Table 9 below shows the response rates, contact rates and cooperation rates for each of the 16 ESS countries; first when using the final outcome distributions derived by means of priority coding, second when using the distributions resulting from most-

recent coding and finally when using the final outcome distributions that the national teams derived themselves and provided to the ESS as part of their technical report (ESS, 2003).

The calculation of outcome rates was guided by the AAPOR standard definitions (AAPOR, 2006); however, it was not possible to directly apply any of the AAPOR outcome rates, since the final case outcomes were defined differently in the ESS. The most notable difference is that the AAPOR rates give guidance on how to treat cases of unknown eligibility, yet, as discussed, the ESS outcome codes do not provide for cases of unknown eligibility. Furthermore, the ESS contact forms contain cases with missing outcome codes. I assume here that these cases were eligible, yet not contacted. The adapted formulas I used to calculate the outcome rates are as follows.[11]

$$\text{Response rate} = \frac{I}{(I + P) + (R + NC + O) + MIS}$$

$$\text{Contact rate} = \frac{(I + P) + (R + O)}{(I + P) + (R + NC + O) + MIS}$$

$$\text{Cooperation rate} = \frac{I}{(I + P) + (R + O)}$$

, where

I	=	Complete interview
P	=	Partial interview
R	=	Refusal and break-off
NC	=	Non-contact
O	=	Other contact
MIS	=	Outcome missing

[11] For case outcomes derived by means of priority or most recent coding the definitions of these outcome categories correspond to those listed in Table 6. For case outcomes that were derived nationally and where the outcome distributions were taken from the technical reports (ESS, 2003), the following definitions apply:

Complete interview	=	Records on the data file (row p)
Partial interview	=	Not defined in the ESS technical report
Refusal and break-off	=	Refusal by respondent (row b); Refusal by proxy (row c)
Non-contact	=	No contacts (row d)
Other contact	=	Language barrier (row e); Respondent mentally or physically unable to cooperate throughout fieldwork period (row f); Respondent unavailable throughout fieldwork period for other reasons (row g)
Outcome missing	=	Number of sample units not accounted for (row x), interviews not approved (row o)

The differences in response, contact and cooperation rates between priority and most-recent coding in Table 9 correspond to what was discussed above. Differences in response rates between the two are marginal, since the differences in number of ineligible cases was small and both coding strategies were adjusted for the number of interviews in the main data file. Also when comparing these two response rates to the response rate calculated with the outcome distributions provided by each country in their technical report, the differences are not considerable. The largest difference is found in Germany, where the response rate is approximately 3.4 percentage points higher when the country distributions are used. Other countries that derived case outcomes leading to somewhat higher response rates were Greece, Italy, Luxembourg, Poland and Spain. Conversely, the outcome codes provided by Slovenia yield a 1 percentage point lower response rate.

Table 9: Outcome rates across countries by coding strategy

	Response rate			**Contact rate**			**Cooperation rate**		
	Priority %	Most-recent %	Country coding %	Priority %	Most-recent %	Country coding %	Priority %	Most-recent %	Country coding %
Austria	60.5	60.5	60.4	89.8	88.8	88.0	67.4	68.1	68.7
Belgium	59.3	59.2	59.2	93.2	90.0	91.9	63.6	65.8	64.4
Finland	73.3	73.2	73.2	98.2	94.1	96.1	74.7	77.8	76.2
Germany	52.3	52.2	56.5	87.6	85.3	90.1	59.7	61.1	61.8
Greece	79.5	79.5	80.0	98.0	97.3	97.2	81.2	81.7	82.2
Hungary	69.9	69.8	69.9	91.3	89.1	93.2	76.6	78.3	75.0
Ireland	64.8	64.2	64.5	90.5	88.6	90.2	71.7	72.5	71.5
Italy	42.5	42.5	43.7	94.6	92.2	95.9	44.9	46.1	45.6
Luxembourg	43.3	43.3	43.9	89.9	88.9	88.5	48.2	48.7	49.6
Netherlands	67.9	67.9	67.9	97.3	92.6	95.4	69.7	73.3	71.1
Norway	65.0	65.0	65.0	96.1	96.1	97.3	67.7	67.7	66.8
Poland	72.2	72.2	73.2	97.8	96.1	97.9	73.8	75.1	74.8
Portugal	69.1	69.1	68.8	96.3	94.8	95.9	71.8	72.9	71.8
Slovenia	71.5	71.5	70.5	94.1	91.3	92.8	76.1	78.3	75.3
Spain	52.5	52.5	54.0	89.7	87.3	88.6	58.5	60.1	60.0
UK	55.2	55.2	55.5	93.7	85.3	93.8	59.0	64.7	59.2

The difference in non-contact and refusal rates between priority and most-recent coded outcomes corresponds to what the analysis of outcome distributions showed previously. Most-recent coding consistently yields a lower contact rate, while priority coding reveals a consistently lower cooperation rate. Interesting is the comparison of

these rates with those calculated by means of the case outcome distributions provided by the countries. The country contact and cooperation rates tend not to correspond to rates calculated with either priority or most-recent coding. For example, in seven out of 16 countries the country contact rate lies somewhere in between the priority and most-recent contact rates (Belgium, Finland, Ireland, the Netherlands, Portugal, Slovenia and Spain). In four out of 16 countries the contact rate estimated with the country data is higher than both the most-recent and the priority contact rate (Germany, Hungary, Italy, Norway). In two countries the country contact rate is lower than the other two rates (Austria and Luxembourg). Only in Poland and the UK is the country contact rate almost the same as the priority coded contact rate, and only in Greece is the country contact rate almost the same as the contact rate calculated with the most-recent case outcomes. The variation in cooperation rates is similar to that of the contact rates.

Note again that each outcome rate (response rate, contact rate, cooperation rate) was calculated with the same formula for each of the two methods of deriving case outcome codes (priority coded and most-recent coded). Differences in rates therefore stem solely from differences in how the final case outcomes were derived.[12] The derivation of final case outcomes therefore plays an important role if one wants to compare contact and cooperation rates, and to a lesser extend also response rates, across countries. For example, while Greece and Norway had quite comparable contact rates according to their own coding strategies (97.2% and 97.3% respectively) the priority coded contact rates in the two countries differed by almost two percentage points (98.0% in Greece and 96.1% in Norway).

I conclude this section with a summary of the findings and some general reflection on what this means for the comparability of response outcomes and outcome rates. The theoretical discussion of suitability of coding strategies for deriving case outcomes has shown that most-recent coding might be easier to apply (especially in paper-based settings), however, priority coding better reflects the extent to which the ultimate goal

[12] Though there is also some scope for countries interpreting the ESS case outcomes in the technical report differently. For example, countries with samples of individuals might treat some of the contact outcomes as contact with the sample unit only (rather than including household contact), yet the scope for this given the ESS outcome categories in the technical report is limited.

of fieldwork, i.e. conducting an interview, has been achieved. Priority coding should therefore be the method of choice for deriving case outcomes.

The way in which a case outcome was derived from a sequence of call outcomes has proven to have an impact of the distribution of case outcomes and the magnitude of outcome rates, particularly on non-contact and refusal outcomes. Especially surveys with extensive refusal conversion efforts are affected by this. Most-recent coding yields more non-contact case outcomes and consequently a lower contact rate. At the same time, fewer refusals and other contact case outcomes are derived. Conversely, priority coding yields fewer non-contact outcomes and consequently a higher contact rate. At the same time, the cooperation rate tends to be lower, due to the larger base of contacted case outcomes.

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" (Couper and de Leeuw, 2003). My analyses have demonstrated that the effect of differential coding on nonresponse rates is marginal. However, it is increasingly acknowledged that nonresponse analyses (within and across countries) have to consider contact and cooperation separately. Differences in coding strategies bring about differences in contact and cooperation rates. If countries differ as to how they derive their final case outcomes and if these case outcomes are used in comparative analyses of contact and cooperation, the conclusions might be misleading. Responding to Couper and de Leeuw's (2003) call for a disclosure of nonresponse sources, my analyses therefore highlight that this disclosure needs to include the coding strategy for deriving case outcomes. Ideally, case outcomes in cross-national nonresponse analyses are derived by means of a uniform priority coding strategy.

Summary and conclusion

This paper set out to investigate three specific aspects of the process of arriving at comparable response outcome data: (1) to detect and conceptualise influences on the

types of response outcomes collected for a survey; (2) to develop a codeframe of response outcomes that will yield outcomes of high comparability across countries and surveys; and (3) to investigate the impact on case outcomes and outcome rates of priority coding and most-recent coding of call outcomes into case outcomes.

The conceptual framework developed in section one identified two types of influences on the available response outcomes. On the one hand, these are the de facto response outcomes influenced by constraints posed by the survey environment. These include manipulable design features (such as mode of interview or the definition of the population), fixed design constraints (such as available sampling frames) and external factors (such as population characteristics). On the other hand, the survey manager or methodologist specifies the type of response outcomes she needs to be collected for evaluating fieldwork efficiency and nonresponse bias and the means by which they are recorded. This includes whether call- or case-level data are collected and stored, how detailed the response outcomes are and whether the contact data are used for case management only or their use for methodological research is already considered at the planning stage.

The cross-national codeframe of response outcomes is described in section two. It is a combined call- and case level codeframe and specifies the final case-level outcomes that need to be derived to achieve comparable outcomes, but also describes how the codeframe may be implemented at the call-level. I distinguish three types of outcomes: (1) in-office case outcomes, which are never implemented in the field yet still determine the final case outcome of a sample unit; (2) final case outcomes, which are equivalent to the final call outcome implemented in the field; and (3) derived case outcomes, which are derived from a sequence of non-final call outcomes from the field. The codeframe is developed to yield a high level of comparability of response outcomes across face-to-face surveys of individuals in households, where one individual per household is interviewed. The hierarchical structure of three levels permits implementation across countries down to level-2 comparability. This enables countries to adapt the codeframe at the lowest level to include country-specific outcome codes.

The final section assesses the impact on outcome distributions and rates of different derivations of final case outcomes from call-level data. Using the contact data of 16 countries that participated in the first round of the ESS, the analyses in this section derived ESS case outcomes by means of priority and most-recent coding. The comparison of outcome distributions showed that priority coding yielded considerably more refusal cases and fewer non-contact cases than most recent coding. Furthermore, a comparison of outcomes from these two strategies with outcome distributions reported in the technical reports of each country showed that the national case outcomes corresponded neither to the priority coded nor to most-recent coded outcomes. Therefore these national distributions of final case outcomes and the resulting contact and cooperation rates can lead to misleading conclusions about cross-national differences in nonresponse.

In the introduction the paper mentions three equivalence criteria for contact data proposed by Blom, Jäckle and Lynn (2008, forthcoming) that the subsequent analysis considers: (1) equivalence of design, (2) equivalence of implementation and (3) equivalence of coding fieldwork outcomes. Some of the aspects of the first equivalence criterion were dealt with in sections one and two of this paper and section three has investigated the third equivalence criterion. What remains is an evaluation of the impact of other differences in the design of contact data (such as the structure of the contact forms or the translation of outcome categories) and of the impact of differences in interviewer training and fieldwork monitoring. These issues can only be investigated under experimental conditions, if at all, and are therefore outside the realm of this paper.

This paper contributes to the academic survey methodology literature as well as offering practical solutions for implementation in the field. Survey managers developing cross-country surveys and striving for comparability of response rates will be interested in section two dealing with the cross-national codeframe of response outcomes. They should however also note the findings from section three, which demonstrate the importance of describing how final case outcomes (i.e. final disposition codes) were derived. I hope that the first and third sections will encourage feedback from the survey methodology community. The relevance of the various influences on response outcomes can only be evaluated by means of their comments.

Likewise I welcome a discussion of the importance of differently coded case outcomes for current research into relationship between nonresponse rates and nonresponse bias.

Appendix

Table A1: Misclassifications of interviews

	Priority coding		**Most-recent coding**	
	Completed interview according to main data file, but not in contact data	Completed interview according to contact data, but not in main data file	Completed interview according to main data file, but not in contact data	Completed interview according to contact data, nut not in main data file
Total number of misclassifications	449	14	532	12
Final case outcomes in contact data of the misclassified interviews				
Interview broken off/incomplete, never completed	58		55	
Interview undefined	204		222	
Refusal by respondent or someone else	38		20	
Appointment / broken appointment	82		72	
Final outcome missing	18		22	
(Any other case outcomes)	49		141	
Total number of cases	31,447	22,027	31,447	22,027

Table A2: Misclassifications of interviews per country

	Priority coding		**Most recent coding**	
	Completed interview according to main data file, but not in contact data	Completed interview according to contact data, but not in main data file	Completed interview according to main data file, but not in contact data	Completed interview according to contact data, but not in main data file
Belgium	33	0	37	0
Germany	25	0	25	0
Finland	5	0	34	0
Hungary	99	0	101	0
Norway	6	0	9	0
Poland	0	0	0	0
Slovenia	150	2	152	1
Austria	0	0	0	0
Spain	8	0	8	0
UK	96	5	130	4
Greece	3	0	3	0
Ireland	1	0	9	0
Italy	0	0	0	0
Luxembourg	16	0	17	0
Netherlands	0	0	0	0
Portugal	7	7	7	7

References

AAPOR (2000). *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys, Second Edition*. Ann Arbor, MI: American Association for Public Opinion Research.

AAPOR (2006). *Standard Definitions of Case Codes and Outcome Rates for Surveys*. American Association for Public Opinion Research.

Billiet, Jaak, Michel Philippens, Rory Fitzgerald and Ineke Stoop (2007). Etimation of Nonresponse Bias in the European Social Survey: Using Information From Reluctant Respondents. *Journal of Official Statistics*, 23(2), pp. 135-62.

Blom, Annelies, Annette Jäckle and Peter Lynn (2008, forthcoming). *Understanding Cross-National Differences in Unit Nonresponse: The Role of Contact Data* in J. Harkness, M. Braun, B. Edwards, T. Johnson, L. Lyberg and P. Ph. Mohler (eds.). *Survey Methods in Multinational, Multiregional, and Multicultural Contexts*. New York: Wiley & Sons.

Blom, Annelies, Peter Lynn and Annette Jäckle (2008). Understanding Cross-National Differences in Unit Nonresponse: The Role of Contact Data. *ISER Working Paper Series*. 2008-01. Institute for Social and Economic Research.

Brick, Michael, Jill Montaquila and Fritz Scheuren (2002). Estimating Residency Rates for Undetermined Telephone Numbers. *Public Opinion Quarterly*, 66(1), pp. 18-39.

Carton, Ann and Annelies Blom (2008). Introduction to the Special Interest Session on Paradata. *International Workshop on Household Survey Nonresponse*. Ljubljana

CASRO (1982). On the Definition of Response Rates. Port Jefferson, NY: Council of American Survey Research Organizations.

Couper, Mick P. (1998). Measuring Survey Quality in a Casic Environment.

Couper, Mick P. and Edith de Leeuw (2003). *Nonresponse in Cross-Cultural and Cross-National Surveys* in J. Harkness and F. J. R. van de Vijver (eds.). *Cross-Cultural Survey Methods*. pp. 157-178 New York: Wiley.

Couper, Mick P. (2008). Personal communication, 4-11-2008.

de Heer, Wim (1999). International Response Trends: Results of an International Survey. *Journal of Official Statistics*, 15(2), pp. 129-42.

ESS (2003). *ESS round 1: Documentation report*. European Social Survey.

Groves, Robert M., Floyd J. Fowler, Mick P. Couper, James M. Lepkowski, Eleanor Singer and Roger Tourangeau (2004). *Survey Methodology*, New Jersey: John Wiley and Sons.

Groves, Robert M. and Steven G. Heeringa (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(3), pp. 439-57.

Hoffmeyer-Zlotnik, Jürgen H. P. and Uwe Warner (2008). Private Household Concepts and Their Operationalisation in National and International Social Surveys. *GESIS Forschungsberichte, Reihe: Survey Methodology*, 1.

Kviz, Frederic J. (1977). Toward a Standard Definition of Response Rate. *Public Opinion Quarterly*, 41(2), pp. 265-67.

Lalor, Thérèse and Tim Wardrop (2007). Data Collection Challenges in Australian Remote Indigenous Communities. *Conference of the European Social Research Association*. 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). *Proceedings of the Survey Research Methods Section*. Alexandria, VA: American Statistical Association.

Lin, I-Fen and Nora Cate Schaeffer (1995). Using Survey Participants to Estimate the Impact of Nonparticipation. *Public Opinion Quarterly*. 59(2), pp. 236-258.

Loosveldt, Geert, Michel Philippens, Ineke Stoop and Jaak Billiet (2003). Refusal Conversion Procedures in the European Social Survey. *International Workshop on Household Survey Nonresponse*. Leuven.

Lynn, Peter (2003). Developing Quality Standards for Cross-National Survey Research: Five Approaches. *International Journal of Social Research Methodology*, 6(4), pp. 323-36.

Lynn, Peter, Roeland Beerten, Johanna Laiho and Jean Martin (2001). *Recommended Standard Final Outcome Categories and Standard Definitions of Responde Rate of Social Surveys*. 2001-23. Essex: ISER.

Lynn, Peter, Roeland Beerten, Johanna Laiho and Jean Martin (2002). Towards Standardisation of Survey Outcome Categories and Response Rate Calculations. *Research in Official Statistics*, 5(1), pp. 61-85.

Lynn, Peter and Paul Clarke (2002). Separating Refusal Bias and Non-Contact Bias: Evidence From UK National Surveys. *The Statistician*, 51(3), pp. 319-33.

Mccarty, Christopher (2003). Differences in Response Using Most Recent Versus Final Dispositions in Telephone Surveys. *Public Opinion Quarterly*, 67, pp. 396-406.

Quaglia, Martina (2007). Some Questions Raised While Piloting a Survey on International Migrations Conducted in Senegal and Europe. *Conference of the European Social Research Association*. Prague.

Smith, Tom (2002). *Developing Nonresponse Standards* in R. M. Groves, D. A. Dillman, J. L. Eltinge and R. J. A. Little (eds.). *Survey Nonresponse*. pp. 27-40 New York: Wiley.

Stoop, Ineke, Silke Devacht, Jaak Billiet, Geert Loosveldt and Michel Philippens (2003). The Development of a Uniform Contact Description form in ESS. *International Workshop on Household Survey Nonresponse*. Leuven.

Stoop, Ineke (2005). *The Hunt for the Last Respondent.* The Hague: SCP.