A COUNTRY-BY-COUNTRY EVALUATION OF THE CONTACT FORM DATA IN THE EUROPEAN SOCIAL SURVEY, ROUND 2 (2004)

Stefaan Pleysier

Leen Vandecasteele

CONTENTS

A. INTRODUCTION

In this report, an overview is given of the main issues to be aware of when using the contact form data of the second round of the European Social Survey. The focus is on the comparability and general data quality of the provided datasets. We start with listing the countries for which contact form data is available, including the sample type and sample size of the contact form data. Consequently a country-by-country overview is given of the availability and comparability of variables, as well as issues relating to data quality and data consistency.

B. COUNTRIES INCLUDED IN THE INTEGRATED CONTACT FORM DATAFILE

In the second round of the European Social Survey, contact form data are available for 23 countries. The countries are presented here in alphabetical order. Next to the sample type, also the number of observations in the contact form file is given.

Contact form data for Turkey and Iceland are not included in the integrated file. The data for Turkey are not available; for Iceland the data are seriously incomplete and therefore not comparable. Furthermore, the five countries flagged with * also show data quality problems with varying levels of seriousness. In the country-by-country overview, we refer to these problems briefly. For a more elaborate discussion of these issues, we refer to Billiet & Pleysier (2007).

1. AUSTRIA CF_AT, household sample, 3672 observations
2. BELGIUM CF_BE, individual sample, 3018 observations
3. CZECH REPUBLIC* CF_CZ, household sample, 4335 observations
4. DENMARK CF_DK, individual sample, 2441 observations

5. ESTONIA CF_EE, individual sample, 2867 observations
6. FINLAND CF_FI, individual sample, 2900 observations
7. FRANCE CF_IN_FR, household sample, 4400 observations
8. GERMANY CF_DE, individual sample, 5868 observations
9. GREECE CF_GR, household sample, 3056 observations
10. HUNGARY CF_IN_HU, individual sample, 2463 observations
11. IRELAND CF_IE_UPDATE, address sample, 3981 observations
12. ITALY CF_IT, address sample, 2610 observations
13. LUXEMBOURG CF_LU, individual sample, 3497 observations
14. NORWAY CF_NO, individual sample, 2750 observations
15. POLAND CF_PL, individual sample, 2399 observations
16. PORTUGAL CF_PT, household sample, 3094 observations
17. SLOVAKIA CF_IN_SK, individual sample, 2500 observations
18. SLOVENIA* CF_SI, individual sample, 2206 observations
19. SPAIN CF_ES, individual sample, 3213 observations
20. SWEDEN CF_SE, individual sample, 3000 observations
21. SWITZERLAND* CF_CH, household sample, 4863 observations
22. UKRAINE* CF_IN_UA, address sample, 2866 observations
23. UNITED KINGDOM* CF_UK, address sample, 4032 observations

C. COUNTRY BY COUNTRY EVALUATION OF CONTACT FORM DATA

1. AUSTRIA

119 variables are present in the original country-provided dataset; the missing three variables (HHSELECT, MULTHUM and NUMHH) are not applicable in Austria. All variable names are conform to the ESS data protocol.

In general no wild or out of range codes are observed (specific issues below).

Specific issues:

OUNTIA1-…	notably high number of appointments made during first visits
REFVIS2	a second refusal is reported at the first contact attempt
REFVIS3	a third refusal is reported at the second contact attempt
HOURV1	1 visit at 6 a.m.
HOURV7	1 visit at 6 a.m.

2. BELGIUM

All 122 variables required in the contact data file are present in the original country-provided dataset, and variable names are conform to the ESS data protocol.

Almost no wild or out of range codes are observed (exception below).

Dataset is relatively consistent concerning item non-response through comparable variables; however, some problems with item non-response in parts of the questionnaire appear (e.g.: OUTNIA1: outcome when there was no interview (visit 1): high number of missing values although low number of interviews at visit 1 (RESULA1).

Additionally, the number of cases in the contact form file does not equal the total number of issued sample units (See: NSD). For more information, see Billiet & Pleysier (2007).

Specific issues:

HOURV1-10 HOURV1-2: visit at 1 a.m.

HOURV2: 1 visit at 1 a.m.; 1 visit at 5 a.m.

3. CZECH REPUBLIC

All 122 variables required in the contact data file are present in the original country-provided dataset, and variable names are conform to the ESS data protocol.

In Czech Republic, 1196 sample addresses were not used; these are not included in the contact form dataset. Caution is required when using the data for aggregate figures. For more information, we refer to Billiet & Pleysier (2007).

In general no wild or out of range codes are observed.

There are some problems of item non-response throughout the questionnaire (e.g.1: DAYV10 counts 1 visit, but HOURV10 has 47 visits; e.g.2: OUTNIA1: outcome when there was no interview (visit 1)': 3153 missing, but only 1681 interviews at visit 1 (RESULA1); e.g.3: reason for refusal is not always given (REFVIS1 and RERSA1_1 have different amount of missing values).

4. DENMARK

All 122 variables required in the contact data file are present in the original country-provided dataset, and variable names are conform to the ESS data protocol.

In general no wild or out of range codes are observed (specific issues below).

Similar problems with item non-response throughout the questionnaire (e.g.1: OUTNIA1: outcome when there was no interview (visit 1)': 1184 missing, but only 240 interviews at visit 1 (RESULA1)).

Specific issues:

REFVIS1 number of visit at which refusal occurred: apparently in one case there have been 11 visits (most variables only count up to ten).

REFVIS2 three second refusals reported at the first contact attempt

RERSA1_1 (first) reason for first refusal: there should be 1878 missings (according to REFVIS1), but there are only 21: 2129 'respondents' are coded '10' ('previous bad experience')?

RERSA2_1 idem: according to REFVIS2 only 5 second refusals occur, however there are 39 (second) reasons offered? According to REFVIS3, no third refusals occur; however, one reason is given for a third refusal (RERSA3_1)

5. ESTONIA

All 122 variables required in the contact data file are present in the original country-provided dataset, and variable names are conform to the ESS data protocol.

More than 10 contact attempts were registered in the original country-provided dataset; the second last and last contact attempt is registered under the variables xxx2ndlast and xxxlast. Also some other variables are enlarged in the original country-provided dataset: REFVIS1-REFVIS5, and COOP1-COOP5. Only REFVIS1-REFVIS3 and COOP1-COOP3 are provided in the integrated contact data file.

Almost no wild or out of range codes are observed (exception below).

The dataset is relatively consistent concerning item non-response through comparable variables; however, some problems with item non-response in parts of the questionnaire appear (e.g.: OUTNIA1: outcome when there was no interview (visit 1)': high number of missing cases although 'low' number of interviews at visit 1 (RESULA1).

Specific issues:

RERSA1_2-5 no codes: only 1 reason for refusal at first, second and third refusal?
RERSA2_2-5
RERSA3_2-5
COOP1-3 no information on the 'estimation of the cooperation rate at REFVIS1-3?

6. FINLAND

118 variables are present in the original country-provided dataset; the missing four variables (HHSELECT, MULTHUM, NHHMEM and NUMHH) are not applicable in Finland. All variable names are conform to the ESS data protocol.

In general no wild or out of range codes are observed (exceptions below).

Similar problems with item non-response in parts of the questionnaire (e.g.1: OUTNIA1: outcome when there was no interview (visit 1)': 1020 missing, but only 67 interviews at visit 1 (RESULA1);

Specific issues:

HOURV1-10 possible 'outliers': HOURV1: 7 'visits' at midnight? Similar problems: HOURV7 (1 visit at 6 a.m.); HOURV7 (1 visit at 2 a.m.?); HOURV9 (1 visit at midnight?);

RERSA1_1 There is some divergence between number of first refusals (REFVIS1) and number of (first) reason for this refusal (difference: 47); same problem does not reappear with second refusal, nor third refusal;

AGER/GENDERR AGER and GENDERR have been asked to all respondents: in most other countries this question is only asked if no interview was achieved.

7. FRANCE

All 122 variables required in the contact data file are present in the original country-provided dataset. There is a slight variation in the variable names for MODEV1-MODEV10 and RERSA11-RERSA15, RERSA21-RERSA25, RERSA31-RERSA3_5. These variable names have been adapted.

8. GERMANY

A number of variables are not available in the original country-provided dataset:

- AGER, GENDERR & TELNUM
- HHSELECT, MULTHUM, NUMHH, NHHMEM are not applicable for Germany, they are not present in the original country-provided dataset.

More than 10 contact attempts were registered in the original country-provided dataset; the second last and last contact attempt is registered under the variables xxx2ndlast and xxxlast. Also some other variables are enlarged in the original country-provided dataset: INTNUM1-INTNUM4 (normally INTNUM3), TOTCINT1-TONTCINT4 (normally TOTCINT1-TOTCINT3); REFVIS1-REFVIS8 (normally REFVIS3). For these variables only the first three attempts are provided.

Specific issues:

HOURV4	visit/contact attempt at 0 a.m.?
RESULA1-4	- Code 0: all contact information is missing because all contact form data is missing (for example because of technical problems with the notebook). - Code 8: there is no contact form data as the interview was found to be fake (it proved to be false in the back-check process) and the contact form data was fake too (fictious, made up by the interviewer).

9. GREECE

All 122 variables required in the contact data file are present in the original country-provided dataset, and variable names are conform to the ESS data protocol.

Almost no wild or out of range codes are observed (exception below).

Dataset is relatively consistent concerning item non-response through comparable variables; however, some problems with item non-response in parts of the questionnaire appear (e.g.: OUTNIA1: outcome when there was no interview (visit 1)': high number of missings although 'low' number of interviews at visit 1 (RESULA1);

Specific issues:

INTNUM	there are 10 cases were a 4th interviewer reissue emerged
TOTCINT2-3	codes '0' appear; these should all be coded missing '.' (not applicable)
HOURV2	1 visit at 0 a.m. (midnight)?; 1 visit at 2 a.m.
OUTINELI	only 1 'ineligible'

10. HUNGARY

118 variables are present in the original country-provided dataset; the missing four variables (HHSELECT, MULTHUM, NHHMEM and NUMHH) are not applicable in Hungary. All variable names are conform to the ESS data protocol.

Specific issues:

HOURV3	1 visit at 6 a.m?

11. IRELAND

All 122 variables required in the contact data file are present in the original country-provided dataset, and variable names are conform to the ESS data protocol.

12. ITALY

All 122 variables required in the contact data file are present in the original country-provided dataset, and variable names are conform to the ESS data protocol.

Specific issues:

HOURV1	851 visits at 0 a.m. (midnight)?
	2 visits at 1 a.m.
	1 visit at 5 a.m.
HOURV2	247 visits at 0 a.m. (midnight)?
HOURV3	143 visits at 0 a.m. (midnight)?
HOURV4	103 visits at 0 a.m. (midnight)?
	1 visit at 1 a.m.
HOURV5	10 visits at 0 a.m. (midnight)?

HOURV6	3 visits at 0 a.m. (midnight)?
HOURV9	1 visit at 0 a.m. (midnight)?

13. LUXEMBOURG

All 122 variables required in the contact data file are present in the original country-provided dataset, and variable names are conform to the ESS data protocol.

Almost no wild or out of range codes are observed (exception below).

Dataset is relatively consistent concerning item non-response through comparable variables; however, some problems with item non-response in parts of the questionnaire appear (e.g.: OUTNIA1: outcome when there was no interview (visit 1)': high number of missings although 'low' number of interviews at visit 1 (RESULA1).

Specific issues:

INTNUM2-3	no observations: no re-issues by other interviewers?
MONV1, 2, 4	interviews in month 4 and 5?
HOURV3	visit at 2 a.m.?

14. NORWAY

A number of variables are not available in the original country-provided dataset:

- HHSELECT, NHHMEM and NUMHH are not applicable for Norway, they are not present in the original country-provided dataset.
- TYPE, LITTER, PHYS and VANDA are missing.

All variable names are conform to the ESS data protocol. Almost no wild or out of range codes are observed (exception below). Dataset is very consistent with respect to item non-response on comparable variables.

Specific issues:

HOURV1-10 HOURV1: visit at 0 a.m. (midnight)?; 2 visits at 1 a.m.
HOURV2: visit at 2 a.m.?
HOURV3-4-5: visit at 1 a.m.?

AGER/GENDERR AGER and GENDERR have been asked to all respondents: cf above with Finland.

15. POLAND

All 122 variables required in the contact data form file are present in the original country-provided dataset, and variable names are conform to the ESS data protocol.

In general no wild or out of range codes are observed (exceptions below).

Problems with item non-response in parts of the questionnaire (e.g.1: OUTNIA1: outcome when there was no interview (visit 1)': 1244 missing, but 'only' 843 interviews at visit 1 (RESULA1);

Specific issues:

HOURV1 possible 'outliers': 1 'visit' at 6 a.m.?

16. PORTUGAL

All 122 variables required in the contact data form file are present in the original country-provided dataset, and variable names are conform to the ESS data protocol.

In general no wild or out of range codes are observed (specific issues below). Some similar problems with item non-response throughout the questionnaire as in the above countries (cf. OUTNIA1: 2620 missing?).

Specific issues:

HOURV1	visit at 1, 2, 3, 4 and 6 a.m.? Presumably 13... 18 hours was meant?
HOURV2	5 visits at 1 a.m.?
HOURV3	3 visits at 1 a.m.?
HOURV4	1 visit at 6 a.m.?
HOURV6	1 visits at 2 a.m.?
HOURV7	1 visits at 5 a.m.?

17. SLOVAKIA

118 variables are present in the original country-provided dataset; the missing four variables (HHSELECT, MULTHUM, NHHMEM and NUMHH) are not applicable in Slovakia. All variable names are conform to the ESS data protocol.

Specific issues:

HOURV2	visit at 6 a.m.
HOURV3	1 visit at 6 a.m.

18. SLOVENIA

All 122 variables required in the contact data form file are present in the original country-provided dataset, and variable names are conform to the ESS data protocol.

The data contains a high level of item-nonresponse, and many wild codes. There is also a discrepancy between the number of realised interviews in the contact form file and the main file. Additionally, the number of cases in the contact form file does not equal the total number of issued sample units (See: NSD). For more information, see Billiet & Pleysier (2007).

Specific issues:

MODEVA1-10 these variables are not available for all visits or contact attempts (see number of missings);

19. SPAIN

All 122 variables required in the contact data form file are present in the original country-provided dataset, and variable names are conform to the ESS data protocol.

More than 10 contact attempts were registered in the original country-provided dataset. 24 sample persons were contacted (attempts) more than 10 times; only 3 more than 15 times, and only 1 ‘respondent’ was contacted more than 17 times (up to 37 attempts). The second last and last contact attempt is registered under the variables xxx2ndlast and xxxlast.

No obvious wild or out of range codes are observed (exception below). Dataset is relatively consistent concerning item non-response through comparable variables.

Specific issues:

RERSA1_1	RERSA variables have unknown code '77'.
GENDERR	code '7' is not a category in the data protocol?

20. SWEDEN

A number of variables are not available in the original country-provided dataset:
- TOTCINT1 to TOTCINT3 is missing ('number of visits the original, second and third interviewer did').
- HHSELECT, NHHMEM and NUMHH are not applicable for Sweden, they are not present in the original country-provided dataset.
- TELNUM, AGER, GENDERR, OUTINELI, TYPE, PHYS, LITTER and VANDA.
- Finally, only one reason for each refusal is recorded (RERSA1_2 to 5, RERSA2_2 to 5 and RERSA3_2 to 5 are missing as well).

More than 10 contact attempts were registered in the original country-provided dataset; the second last and last contact attempt is registered under the variables XXX2NDLAST and XXXLAST.

Some wild or out of range codes are observed (see below).

Other problems concern item non-response throughout the questionnaire (e.g.1: OUTNIA1: outcome when there was no interview (visit 1)': 1241 missing, but no interview realized at visit 1 (RESULA1); idem OUTNIA2…);

Specific issues:

HOURV2 possible ‘outliers’: 1 ‘visit’ at 2 a.m.; 1 visit at 6 a.m.? Similar problems: HOURV5 (2 visits at 6 a.m.?);

MODEVA… Unknown categories with codes 5 and 6 in original country-provided data-file.

21. SWITZERLAND

A number of variables are not available in the original country-provided dataset:
- HHSELECT, MULTHUM and NUMHH are not applicable for Switzerland; they are not present in the original country-provided dataset.
- Also TYPE and DAYV10 are missing.

More than 10 contact attempts were registered in the original country-provided dataset. 313 were contacted (attempts) more than 10 times; 142 more than 20 times, 70 more than 30 times, 29 more than 40 times, 7 more than 50 times, and 1 up to 59 contact attempts). The second last and last contact attempt is registered under the variables xxx2ndlast and xxxlast.

Also some other variables are enlarged: INTNUM1-INTNUM12 (normally INTNUM3), TOTCINT1-TONTCINT9, and TOTCIN_1-TOTCIN_3 for 10th to 12th visit (normally TOTCINT1-TOTCINT3), REFVIS1-REFVIS8 (normally REFVIS3), RERSA1_1 – RERSA8_1 (but only 3 in stead of 5 possible reasons for refusal), and COOP1-COOP8 (normally COOP3). Only INTNUM1-INTNUM3, TOTCINT1-TOTCINT3, REFVIS1-REFVIS3N RERSA1_1 – RERSA3_1 and COOP1-COOP3 provided in the integrated contact data file.

Some coding inconsistencies can be found in the data, for more information see Billiet & Pleysier (2007).

Some possible problems with item non-response in parts of the questionnaire can be detected (e.g.: OUTNIA: from OUTNIA37 it seems (not necessary the case) as only one potential respondent is followed; there are however about 30 respondents contacted more than 40 times).

22. UKRAINE

The data for Ukraine have to be used with caution because of serious data quality problems, see Billiet & Pleysier (2007):

- Coding inconsistencies
- Loss of 184 records in contact form data
- Identification numbers between contact form dataset and main dataset do not match

Specific issues:

HOURV1	14 visits at 1 a.m.; 6 visits at 2 a.m.; 1 visit at 5 a.m.
HOURV2	4 visits at 1 a.m.

23. UNITED KINGDOM

All 122 variables required in the contact data form file are present in the original country-provided dataset, and variable names are conform to the ESS data protocol.

The data for United Kingdom have to be used with caution because of serious data quality problems, see Billiet & Pleysier (2007):

- Identification numbers between contact form dataset and main dataset do not match
- The contact form data is incomplete

In general no wild or out of range codes are observed (specific issues below).

Item nonresponse: missing values with MINV1-10 variables do not always correspond with the DATE and HOUR variables.

D. FURTHER READING

Billiet, J. and S. Pleysier (2007). Response Based Quality Assessment in the ESS - Round 2. An update for 26 countries. Leuven, K.U. Leuven - Center of Sociological Research**:** 64.

NSD ESS2-2004 Documentation Report., The ESS Data Archive.