

Some technical details on the Austrian Generations and Gender Survey Wave 2.

Research report 36

Isabella Buber

Wittgenstein Centre for Demography and Global Human Capital
(IIASA, VID/ÖAW, WU), Vienna Institute of Demography/Austrian
Academy of Sciences

March 2014

INSTITUT FÜR DEMOGRAPHIE
ÖSTERREICHISCHE AKADEMIE DER WISSENSCHAFTEN

Forschungsbericht Nr. 36

Die Arbeit unterliegt ausschließlich der Verantwortung der beiden herausgebenden Institute und wurde der phil.-hist. Klasse nicht vorgelegt.

Medieninhaber und Herausgeber:

Institut für Demographie
Vienna Institute of Demography
Österreichische Akademie der Wissenschaften
A-1040 Wien, Wohllebengasse 12-14
Tel.: +43-1/51581/7702
Fax: +43-1/51581/7730
E-Mail: vid@oeaw.ac.at
Homepage: www.oeaw.ac.at/vid

Vertrieb:

Verlag der Österreichischen Akademie der Wissenschaften
A-1011 Wien, Postgasse 7, Postfach 471
Tel.: +43-1/51581/3401-3406
Fax: +43-1/51581/400
E-Mail: verlag@oeaw.ac.at
Homepage: <http://verlag.oeaw.ac.at/>

Some technical details on the Austrian Generations and Gender Survey Wave 2

Isabella Buber

March 2014

Sampling, data collection, fieldwork maintenance and raw data cleaning was carried out by Statistics Austria. Data for the second wave of the Austrian Generations and Gender Survey (GGS) were delivered by Statistics Austria via a file called dg5_AT_wave2.dta (dg5 is the abbreviation for data generation 5). This file is not harmonized, names of variables are those generated by Statistics Austria and deviate from the final variable names in harmonized data. Current variable names reflect the number of the corresponding variable.

The prefix “w2_” indicates that the variable belongs to wave 2. For example, “w2_f120” is the variable assigned to question F120 in the questionnaire of the second wave of the Austrian GGS. The questionnaire, including loops and detailed remarks is available online (Neuwirth and Buber 2013). Moreover, the file includes numerous variables generated by Statistics Austria which were used for internal checks and validation. These variables were not excluded, it is up to the user to drop them if required.

The second wave of the Austrian GGS includes 4,729 records. The majority are panel respondents, refresher constitute a minority (Table 1). The variable “w2_folgebefragung” allows to identify panel respondents and refreshers.

Table 1: Austrian GGS wave 2 data

	Number of respondents
Panel respondents	3,907
Refreshers	822
Total	4,729

1. Corrections

Statistics Austria carried out first data validation. Further checks for panel respondents were carried out at VID. The focus of these checks was on children. Therefore the number of children reported in wave 1 was compared with the number of children in wave 2. Information on children (month and year of birth, type of children like own child/stepchild/adopted child) given in wave 1 and wave 2 were studied in detail and revealed different types of mismatch. If information given in wave 1 and wave 2 differed, and further checks lead to the conclusion that children were most probably the same (i.e. gender, status) as a general rule the values in wave

2 were replaced by those given in wave 1.¹ Variables indicating that changes have been made were generated and are described in Table 2. Values 0 indicate that no changes have been made, value 1 indicates that changes have been made in W2.

Table 2: Different types of corrections of the wave 2 data

Variable	Label
w2_correct1	No panel
w2_correct2	ID swap
w2_correct3	Year or month of birth of respondent was corrected in W2
w2_correct4	Dead child in W2 already reported in W1
w2_correct5	Household or non-household or child was missing in W2
w2_correct6	Discrepancies for children were corrected in W2
w2_correct7	Partner's birthdate was corrected in W2 according to W1
w2_correct8	W1 has to be corrected, dead child reported in W2 but not in W1
w2_correct9	W1 has to be corrected, inconsistencies for children in W1 and W2

It turned out that in 5 cases, a different person was interviewed in wave 2 as in wave 1. These records are in fact refreshers and not panel respondents (Table 3). Two IDs were swapped and IDs were corrected to allow match of same persons.

In a total of 14 records respondent's month and/or year of birth slightly differed in wave 1 and wave 2, all other information on the children being identical (i.e. sex of child, type of child). For these records, respondent's year and/or month of childbirth in wave 2 were replaced by the values given in wave 1.

Reporting deceased children constituted a source of inconsistencies. In both waves, respondents were asked about possible deceased children. In wave 2, interviewers were instructed to record children who died between wave 1 and wave 2 only. Nevertheless, 8 records include deceased children who had died before wave 1 interview and who were already coded in wave 1 (gender, month and year of birth and death in both waves were identical). Therefore, these deceased children were dropped in wave 2. Otherwise, dead children would have been counted twice when generating the total number of children ever born.

Comparing the parity reported in wave 1 and wave 2 revealed that some respondents had higher parity in wave 1 than in wave 2. For these cases, information on children was compared case by case (gender, date of birth, type of child). It turned out that 52 respondents either did not report children who have left the household between wave 1 and wave 2, or did not report in wave 2 children not living in the household at wave 1. After detailed checks, corresponding variables in wave 2 were corrected according to the information given in wave 1. Children not mentioned in wave 2 were included as children not living in the household.

Further checks revealed for 15 interviewees discrepancies for children given in both waves. For example, year and/or month of birth or status of the child were different (e.g. own child in wave

¹ Another possibility would be to correct wave 1. Since we want to keep corrections of wave 1 data to a minimum, we decided to replace in case of inconsistencies wave 2 information by wave 1 information.

1, stepchild in wave 2). For the mentioned 15 records the corresponding variable values were changed according to information given in wave 1.

Besides detailed checks on children, basic checks on the information on partners given in wave 1 and wave 2 were carried out. An indicator was generated indicating – for those with a partner in both waves – if the partners in wave 1 and wave 2 were the same. Therefore birthdate of the partners recorded in both waves were compared. Minor discrepancies (like different month of birth of the partner, the other information being identical in both waves) for 5 records were found and corrected.

Comparing wave 1 and W2 revealed that for several records wave 1 data have to be corrected. It turned out that 6 respondents reported in wave 2 deceased children, who had died before wave 1 interview and who were not reported in wave 1. Moreover, for further 72 records detailed comparison of information provided in waves 1 and 2 lead to the conclusion that wave 1 data have to be corrected according to the information given in wave 2. For these records, wave 1 data have to be corrected. The current available harmonized data on the Austrian GGS wave 1 do not include these corrections, users have to correct wave 1 data individually.

Table 3: Number of corrections of the wave 2 data

	Total
No panel	5
ID swap	2
Year or month of birth of respondent was corrected in W2	14
Dead child in W2 already reported in W1	8
Household or non-household or children was missing in W2	52
Discrepancies for children were corrected in W2	15
Partner's birthdate was corrected in W2 according to W1	5
W1 has to be corrected, dead child reported in W2 but not in W1	6
W1 has to be corrected, inconsistencies for children in W1 and W2	72

The corrections mentioned above will be included in the final harmonized wave 2 data which will be made available to the scientific community in 2014/15. Currently, corrected wave 2 data are available to researchers at the Vienna Institute of Demography (VID) and the Austrian Institute for Family Studies (OIF).

Most probably, there will be further discrepancies between information given in wave 1 and in wave 2. It is not planned to correct discrepancies on partnership histories or employment histories. It is up to the data user to either correct or to drop records with inconsistencies.

2. Generated variables

I generated month and year of birth of all natural children, distinguishing between children living in the household, children not living in the household and dead children reported at wave 2. Moreover, I generated the number of children born between wave 1 and wave 2 (w2_newchild) - a valuable variable when analyzing for example the realization of fertility intentions. As mentioned earlier, deceased children at wave 1 were not coded in wave 2, to avoid double counting. For panel respondents, the birthdates of these children have been extracted from wave 1 and merged with wave 2 data. The corresponding variable names are “w1_dead1m”, “w1_dead1y” through “w1_dead3m” and “w1_dead3y”. Table 4 gives the names of these generated variables.

Table 4: Generated variables for birth histories

Variable	Description
w2_child1m	Month of birth of own child 1 living in the household
w2_child1y	Year of birth of own child 1 living in the household
w2_child2m	Month of birth of own child 2 living in the household
w2_child2y	Year of birth of own child 2 living in the household
w2_child3m	Month of birth of own child 3 living in the household
w2_child3y	Year of birth of own child 3 living in the household
w2_child4m	Month of birth of own child 4 living in the household
w2_child4y	Year of birth of own child 4 living in the household
w2_child5m	Month of birth of own child 5 living in the household
w2_child5y	Year of birth of own child 5 living in the household
w2_child6m	Month of birth of own child 6 living in the household
w2_child6y	Year of birth of own child 6 living in the household
w2_child7m	Month of birth of own child 7 living in the household
w2_child7y	Year of birth of own child 7 living in the household
w2_child1m_nh	Month of birth of own child 1 not living in the household
w2_child1y_nh	Year of birth of own child 1 not living in the household
w2_child2m_nh	Month of birth of own child 2 not living in the household
w2_child2y_nh	Year of birth of own child 2 not living in the household
w2_child3m_nh	Month of birth of own child 3 not living in the household
w2_child3y_nh	Year of birth of own child 3 not living in the household
w2_child4m_nh	Month of birth of own child 4 not living in the household
w2_child4y_nh	Year of birth of own child 4 not living in the household
w1_dead1m	Month of birth of own deceased child 1 as reported in wave 1
w1_dead1y	Year of birth of own deceased child 1 as reported in wave 1
w1_dead2m	Month of birth of own deceased child 2 as reported in wave 1
w1_dead2y	Year of birth of own deceased child 2 as reported in wave 1
w1_dead3m	Month of birth of own deceased child 3 as reported in wave 1
w1_dead3y	Year of birth of own deceased child 3 as reported in wave 1
w1_parity	Parity at wave 1
w2_newchild	Number of children born between W1 and W2

Wave 2 data include several helpful variables generated by Statistics Austria on children, including parity, the number of children living in the household, the number of children not

living in the household, on partner, parents and health². Moreover, I generated a variable indicating if respondents were living at wave 1 and wave 2 with the same partner (w2_same_partner).

Tables below give an overview of these variables. Again, prefix “w2_” indicated wave 2 data. In general, the user has to keep in mind that value -3 stand for „Filter“, value -2 for “don’t know” and value -1 for “Refusal”.

Table 5: Generated variables on number of children in the household

Variable	Label	Description
w2_anz_bez_eq_2	anz. kinder leiblich, jetziger partner, im hh	Number of natural children living in the household with current partner
w2_anz_bez_eq_3	anz. kinder leiblich, früherer partner, im hh	Number of natural children living in the household with previous partner(s)
w2_anz_bez_eq_4	anz. stiefkinder, im hh	Number of stepchildren living in the household
w2_anz_bez_eq_5	anz. adoptivkinder, im hh	Number of adopted children living in the household
w2_anz_bez_eq_6	anz. pflegekinder, im hh	Number of foster children living in the household
w2_k_anz_imhh	anzahl kinder im hh	Number of children living in the household
w2_k_anz_nleibl_imhh	anzahl nicht leibl. Kinder im hh	Number of non-natural children living in the household

Table 6: Generated variables on number of children not in the household

Variable	Label	Description
w2_anz_kindl_nhh	anz. kinder leibl, n. im hh	Number of natural children not living in the household
w2_anz_kinda_nhh	anz. adoptivkinder, n. im hh	Number of adopted children not living in the household
w2_anz_kindp_nhh	anz. pflegekinder, n. im hh	Number of foster children not living in the household
w2_anz_kind_nhh	anz. kinder n. im hh	Number of children not living in the household (including natural, adopted or foster children).
w2_anz_kind_tod	anz. gestorbene kinder	Number of dead children
w2_anz_kinds_nhh	anz. stiefkinder, n. im hh	Number of stepchildren (of current partner) not living in the household
w2_k_anz_nimhh	anzahl kinder nicht im hh	Number of children not living in the household

² The same variables have been generated for wave 1 and are available in the Austrian country file, not in the international harmonized file.

Table 7: Generated variables on age of youngest/oldest child

Variable	Label	Description
w2_k_alter_jk	alter jüngstes kind	Age of youngest child
w2_k_alter_jk_imhh	alter jüngstes kind im hh	Age of youngest child living in the household
w2_k_alter_ak	alter ältestes kind	Age of oldest child
w2_k_alter_ak_imhh	alter ältestes kind im hh	Age of oldest child living in the household

Table 8: Generated variables on parents

Variable	Label	Description
w2_v_alter_imhh	alter vater im hh	Age of father living in the household
w2_v_imhh	vater im hh	Father living in the household
w2_v_bes	beschäftigung d. vaters	Employment status of father
w2_m_alter_imhh	alter mutter im hh	Age of mother living in the household
w2_m_imhh	mutter im hh	Mother living in the household
w2_m_bes	beschäftigung d. mutter	Employment status of mother

Table 9: Generated variables on living arrangement/household composition

Variable	Label	Description
w2_lebensform	lebensform	Living arrangement (detailed information on household composition)
w2_famstand	familienstand	Family status (de jure)

Table 10: Generated variables on partner

Variable	Label	Description
w2_partner	partner vorhanden	Partner existing
w2_p_imhh	partner im hh	Partner living in the household
w2_p_bes	beschäftigung des partners	Employment status of partner
w2_p_geschl	geschlecht des partners	Gender of partner
w2_p_alter	alter partner	Age of partner
w2_p_alter_imhh	alter partner im hh	Age of partner living in the household
w2_same_partner	Same partner generated by Isabella	Indicator whether partners in W1 and W2 are the same or different

Table 11: Generated variables on limitations in activities of daily living (ADL)

Variable	Label	Description
w2_beh_b	befragter beeinträchtigt	Respondent is limited in ADL
w2_beh_p	partner beeinträchtigt	Partner is limited in ADL
w2_beh_v	leiblicher vater beeinträchtigt	Father is limited in ADL
w2_beh_m	leibliche mutter beeinträchtigt	Mother is limited in ADL

Table 12: Generated variables on region

Variable	Label	Description
w2_nuts1		Nuts1
w2_nuts2		Nuts2
w2_nuts3		Nuts3
w2_nuts4		Nuts4
w2_agrarq	agrarquote	Agrarian quota
w2_urban	eurostat-urbanisierungsgrad	Regional typology according to Eurostat

Table 13: Generated variables on education

Variable	Label	Description
w2_artab_b	art der höchsten abgeschlossenen ausbildung	Typ of respondent's highest educational attainment
w2_xartab_b	höchste abgeschlossene bildung	Respondent's highest educational level
w2_hatlevel_b	höchste abgeschlossene schulbildung	Respondent's highest completed school typ
w2_hatfield_b	ausbildungsbereich der höchsten abgeschlossenen schulbildung	Respondent's field of training of highest school typ
w2_artab_p	art der höchsten abgeschlossenen ausbildung	Typ of partner's highest educational attainment
w2_xartab_p	höchste abgeschlossene bildung	Partner's highest educational level
w2_hatlevel_p	höchste abgeschlossene schulbildung	Partner's highest completed school typ
w2_hatfield_p	ausbildungsbereich der höchsten abgeschlossenen schulbildung	Partner's field of training of highest school typ

3. Weights

Weight “w2_weight” is a cross-sectional weight adjusting for age, gender, employment status, country of birth and living arrangements. Furthermore, weights for the female sample adjust for the cohort-specific parity distribution (for 5-year age-cohorts, based on Geburtenbarometer). The same characteristics were taken into account for generating weight for wave 1 (Buber 2010). These weights are poststratification weights.

4. Merging wave 1 and wave 2

For merging wave 1 and wave 2 use the indicator “Ordnungs”.

5. Final data file of the Austrian GGS wave 2

Corrected and renamed data of wave 2 are stored in the file

“dg5_AT_wave2_10032014.dta”.

These data include all corrections carried out by 10 March 2014.

If you find further inconsistencies, please report them to Isabella.buber@oeaw.ac.at

Acknowledgements

I want to thank the researchers at the Vienna Institute of Demography (VID) and at the Austrian Institute for Family Studies (OIF) for valuable discussions during the data validation process. A special thank to Caroline Berghammer who provided STATA code for generating birth histories.

References

- Buber, I. (2010). *Parity-specific weights for the Austrian Generations and Gender Survey*. VID Working Paper 4/2010: Vienna Institute of Demography.
- Neuwirth, N., & Buber, I. (2013). *Familienentwicklung in Österreich. Generationen und Geschlechterrollen. Hauptfragebogen für Welle 2*. Forschungsbericht Nr. 35, Vienna: Vienna Institute of Demography.