SOEP The German Socio-Economic Panel study

538

SOEP Survey Papers

Series D – Variable Descriptions and Coding

SOEP — The German Socio-Economic Panel study at DIW Berlin 2018

SOEP-Core v33.1 – MIGSPELL and REFUGSPELL: The Migration-Biographies of Samples M1/M2 and M3/M4

Klaudia Erhardt and SOEP Group

DIW SOEP

Running since 1984, the German Socio-Economic Panel study (SOEP) is a wide-ranging representative longitudinal study of private households, located at the German Institute for Economic Research, DIW Berlin.

The aim of the SOEP Survey Papers Series is to thoroughly document the survey's data collection and data processing. The SOEP Survey Papers is comprised of the following series:
Series A – Survey Instruments (Erhebungsinstrumente)
Series B – Survey Reports (Methodenberichte)
Series C – Data Documentation (Datendokumentationen)
Series D – Variable Descriptions and Coding
Series E – SOEPmonitors
Series F – SOEP Newsletters
Series G – General Issues and Teaching Materials

The SOEP Survey Papers are available at **http://www.diw.de/soepsurveypapers**

Please cite this paper as follows:

Klaudia Erhardt and SOEP Group. 2018. SOEP-Core v33.1 – MIGSPELL and REFUGSPELL: The Migration-Biographies of Samples M1/M2 and M3/M4. SOEP Survey Papers 538: Series D. Berlin: DIW/SOEP

ISSN: 2193-5580 (online)
DIW Berlin
German Socio-Economic Panel (SOEP)
Mohrenstr. 58
10117 Berlin
Germany
soeppapers@diw.de

SOEP-Core v33.1 – MIGSPELL and REFUGSPELL: The Migration-Biographies of Samples M1 / M2 and M3 / M4

Klaudia Erhardt and SOEP Group

Contents

MIGSPELL and REFUGSPELL: The Migration-Biographies of Samples M1/M2 and M3/M4

Klaudia Erhardt

1 MIGSPELL and REFUGSPELL - why two different data files on migration biographies?

Starting with wave bg (2016), 2 new refugees-samples, M3 and M4, are included in the SOEP. For a detailed description of these samples see M. Kroh et al. (2017): Sampling, Nonresponse, and Integrated Weighting of the 2016 IAB-BAMF-SOEP Survey of Refugees (SOEP Survey Paper 477)[1]. For a general overview of the IAB-BAMF-SOEP Survey see the project page on the DIW website[2].

As with the migration samples M1 and M2, the migration biography up to the last move to Germany is collected during the first wave of each respondent of samples M3/M4. However, the related questions for samples M3/M4 differ significantly from the questions for samples M1/M2, which are captured in file MIGSPELL. Thus, we decided to generate a second file, named REFUGSPELL, to hold the migration biographies of the IAB-BAMF-SOEP Survey of refugees.

Technically, the structure of both files allows for appending to each other. However, we prefer not to provide an integrated spell data set for both migrants and refugees, in order to make it explicitly clear that there are major semantic differences that must be carefully examined before combining the migration biographies into a single data analysing procedure.

Table 1 shows the intersection of both data sets, MIGSPELL and REFUGSPELL, in terms of the variables. At first glance, both files appear to have many variables in common. However, apart from technical variables, the concordance is limited mainly to "When?" and "Where?" The highlighted area of the table marks those variables without good concordance. These describe under what terms a move has taken place and represent the main semantic content of the spell data besides the variables designating "When" and "Where". Despite the general lack of semantic concordance, single characteristics of these variables might be comparable (see the detailed description in sections 2 and 3).

There are further differences between MIGSPELL and REFUGSPELL that must be noted when conducting data analyses, even if it appears to be compatible data. For details, please consult sections 3.1, 3.5, and 4.

[1] http://www.diw.de/documents/publikationen/73/diw_01.c.572346.de/diw_ssp0477.pdf
[2] https://www.diw.de/en/diw_01.c.538695.en

On the other hand, there are quite a few common features. Notwithstanding these, both files are fully documented in two separate chapters. This entails some redundancies, but allows for reading only those parts of the documentation that are relevant if working with only one of the two files.

Table 1: Intersection of MIGSPELL and REFUGSPELL in terms of their variables

Variable	MIGSPELL	REFUGSPELL	Note
hhnr	x	x	ID
persnr	x	x	ID
sample1	x	x	Sample characteristic
stype	x	x	Technical variable, generated
mignr	x	x	Technical variable, generated
nspells	x	x	Technical variable, generated
migfilter	x	x	Redundant to bcountry
country	x	x	"Where?"
starty	x	x	Time specification
startmo	x	x	Time specification
starty_imp	x	x	Time specification, generated
startmo_imp	x	x	Time specification, generated
start	x	x	Time specification, generated
end	x	x	Time specification, generated
staytime	x	x	Time specification, generated
status1	x	-	
status2	x	-	
ostatus	x	-	
rexit	-	x	
jobpr	x	-	
lfgroup	x	-	
move	x	x	Redundant to country, bcountry
nmtype	x	x	Equivalent to move of next spell
tcountry	x	x	Equivalent to country of next spell
bcountry	x	x	Time-invariant data
lastmig	x	x	Time-invariant data
lastmig_imp	x	x	Time-invariant data, generated
lastmig_refug	-	x	Time-invariant data, generated from question outside of the migration biography
f_starty_imp	x	x	Technical variable, generated
f_startmo_imp	x	x	Technical variable, generated
f_birthdate_corr	x	x	Technical variable, generated
f_bio_corr	-	x	Technical variable, generated

Variable	MIGSPELL	REFUGSPELL	Note
f_short	-	x	Technical variable, generated
f_bcountry_iffy	-	x	Technical variable, generated
intyear	x	x	Time-invariant data
censor	x	x	Technical variable, generated
censor_imp	x	x	Technical variable, generated

2 MIGSPELL v33

2.1 Introduction to the release v33 of MIGSPELL

In the previous release, v32, of MIGSPELL the integration of the partly disparate migration biography data of waves bd to bf had been carried out. For a detailed description of the integration, including a minute synopsis showing the mapping of the single values of every variable involved onto the MIGSPELL variables from waves bd to bf, see section 5.

In the present wave, bg, no further changes to the migration-biography related questions have been implemented, meaning that the bg-migration-biography conforms perfectly to the bf-migration-biography. Therefore, the v33 MIGSPELL release at hand has exactly the same structure as the previous release, and the mapping of the variables from file bgp_mig to MIGSPELL is the same as from file bfp_mig, apart from different variable names.

In v33, 129 new respondents who moved into the M1 and M2-households in the last year are included. The v33 release of MIGSPELL contains data from 7,157 respondents and a total of 15,640 spells.

Table 2: MIGSPELL respondents by germanborn/abroadborn and interview year

	Germanborn	abroadborn	Total
2013	1,254	3,710	4,964
2014	98	177	275
2015	166	1,623	1,789
2016	33	96	129
Total	1,551	5,606	7,157

Source data: MIGSPELL v33.1

Table 2 shows that most migration biographies stem from the 2013 and 2015 surveys (M1 and M2), while in 2014 and 2016 only persons who had moved into the sample households were added.

In the following sections, the variables of MIGSPELL and their generation are described in detail.

2.2 Summary description of MIGSPELL

MIGSPELL is derived from the migration biographies, which are collected from each new respondent to the IAB-SOEP Migration Survey, samples M1 and M2.[3] The moves of foreign-born migrants, as well as the stays abroad of German-born respondents, were captured through a complex loop structure within the questionnaire, with the number of loops limited to 15.

Pages 8 and 9 of SOEP Survey Paper 261[4] show two flow charts of the migration biography questions of the IAB-SOEP migration sample. Because compiling such flow charts is very labor-consuming, we provide it only for the first wave of sample M1. The questions regarding the stays abroad of foreign-born and German-born respondents remained unchanged for samples M1 and M2 over all waves. The questions concerning the moves to Germany of foreign-born respondents did change between waves bd and bf. These changes mainly affected the response alternatives (and subsequently, minor filter paths), but not the whole structure as such. Therefore, the flow charts are still helpful for all the waves to date.

The $$p_mig files of the SOEP-distribution hold these data in "wide" format: A distinct set of variables for each potential loop is laid out. To create MIGSPELL, the original data has been transformed into spell format.[5] Each migration that actually took place is represented by a spell.[6] Additionally, a spell is generated for the period from birth to first move (if there has been any), or from birth to interview date (if the respondent has not moved). As a result, all respondents in the IAB-SOEP migration samples are represented in the MIGSPELL dataset, not just those who actually had moved to another country.

The maximum number of spells of a person is 31, because a move to Germany and a move to another country was captured within the same loop cycle, and the spell for the time between birth and first move adds to the number of surveyed moves. However, the distribution of the number of moves is extremely skewed, with only 6 persons having done more than 10 moves, and only one person reaching the limit of 15 loops (see Table 3).

[3] See http://www.diw.de/en/diw_01.c.485464.en for a description and additional documentation of the IAB-SOEP migration samples M1 and M2.

[4] Klaudia Erhardt (2015): Flowcharts for the Integrated Individual-Biography Questionnaire of the IAB-SOEP Migration Sample 2013. SOEP Survey Papers 261, https://www.diw.de/documents/publikationen/73/diw_01.c.570832.de/diw_ssp0261.pdf

[5] For details on the transformation from wide data to spell data, see: Klaudia Erhardt (2014): How to Generate Spell Data from Data in “Wide” Format. SOEP Survey Papers 228: Series G. Berlin: DIW/SOEP, http://www.diw.de/documents/publikationen/73/diw_01.c.570742.de/diw_ssp0228.pdf

[6] To put it exactly, only stays that were a minimum length of three months were supposed to be captured, a rule that was generally observed. Only 1 % of the reported stays are shorter.

Table 3: Number of spells of persons (immigrants only)

nspells -- no of spells by no of persons (group: born outside of Germany)

		Freq.	Percent	Valid	Cum.
Valid	2	4521	80.65	80.65	80.65
	3	372	6.64	6.64	87.28
	4	503	8.97	8.97	96.25
	5	73	1.30	1.30	97.56
	6	79	1.41	1.41	98.97
	7	18	0.32	0.32	99.29
	8	18	0.32	0.32	99.61
	9	8	0.14	0.14	99.75
	10	8	0.14	0.14	99.89
	12	2	0.04	0.04	99.93
	13	1	0.02	0.02	99.95
	14	2	0.04	0.04	99.98
	31	1	0.02	0.02	100.00
	Total	5606	100.00	100.00	

Source data: MIGSPELL v33.1

2.3 Synopsis of the variables in MIGSPELL (systematic order)

Table 4: Variables in MIGSPELL

Variable	Variable label	Description
Identifiers		
hhnr	Original Household Number	Identifier throughout the SOEP
persnr	Never Changing Person ID	Identifier throughout the SOEP
mignr	Running no. of move	MIGPELL-specific identifier. The combination of persnr and mignr identifies a spell. Runs from 0 because first spell is not a move but episode from birth to first move or interview date, respectively.
sample1	Sample Indicator	Sample indicator throughout the SOEP
Time-invariant Characteristics		
sample1	Sample Indicator	
intyear	Interview year	Year when migration biography was surveyed
migfilter	German-born/abroad-born	
bcountry	Country of birth	
lastmig	Year of last move to Germany	= starty of the last spell of a person (only foreign-borns)
lastmig_imp	Year of last move to Germany (imputed version)	= starty_imp of the last spell of a person (only foreign-borns)

Time-variant Characteristics		
starty	Start year	Start year of an episode. Birth year in first spell of a respondent.
startmo	Start month	Start month of an episode. Birth month in first spell of a respondent.
starty_imp	Start year imputed	Same as starty, with imputed missing values and birth year updated with birth year from PPFAD
startmo_imp	Start month imputed	Same as startmo, with imputed missing values and birth month updated with birth month from PPFAD
start	Start (months from jan 1900)	generated from starty_imp and startmo_imp
end	End (months from jan 1900)	= start of next spell - 1
stype	Spell type	Move to Germany / move abroad
country	Country of the stay	
status1	Legal background of entry	source variables see Table 5 to Table 7.
status2	Status at entry	source variables see Table 5 to Table 7, page 14.
ostatus	Status at entry: unstandard. answers	derived from open answers of level "other" of status2, see Table 7
jobpr	Job agreement at entry	source variables see Table 5 to Table 7.
lfgroup	Labor force group	source variables see Table 5 to Table 7.
move	Type of move	to Germany / back to birth country / to another country
nmtype	Type of next move	to Germany / back to birth country / to another country
tcountry	Target country of next move	
staytime	Duration of stay (months)	end - start + 1 (generated from imputed date variables)
Technical Variables		
censor	Censor	Censor Variable generated from original dates
censor_imp	Censor imput. version	Censor Variable generated from imputed dates
nspells	Number of spells	Number of moves of a respondent + 1 (= mignr(max) + 1)
f_birth-date_corr	Flag: birth-date updated with birth-date from PPFAD	= 1 if starty_imp and/or startmo_imp of first spell of a person is updated with PPFAD
f_starty_imp	Flag: starty imputed	= 1 if missing starty has been imputed
f_startmo_imp	Flag: startmo imputed	= 1 if missing startmo has been imputed

2.4 Levels and value labels of the variables in MIGSPELL (alphabetical order)

bcountry Birth country

Coding according to SOEP-convention.

Source variables see Table 5 and Table 6.

censor Censor Marker

Generated variable, based on the original start variables

0 Uncensored
1 Right censored
2 Right miss. censored
3 Left censored
4 Left and right censored
5 Left and right miss. censored
6 Left miss. censored
7 Left miss. and right censored
8 Left miss. and right miss. censored

censor_imp Censor Marker, imputed version.

Generated variable, based on the imputed start variables. Same levels as censor

country Country of the stay.

Coding according to SOEP-convention, = tcountry of preceding spell

end End date of a spell as months from jan 1900

Generated variable = start of the next spell - 1

In last spell, end is generated from interview date (welle in $$p_mig for waves bd - bf, syear in wave bg, and bdpm_pmonin, $$pm_monin for waves be and bf, bgpmmonin)

f_birth-date_corr Flag variable

= 1 if starty and/or startmo of first spell was updated with birth-date from PPFAD

f_startmo_imp Flag variable

= 1 if missing value in startmo was replaced by imputed value

f_starty_imp Flag variable

= 1 if missing value in starty was replaced by imputed value

intyear Year when the migration biography was surveyed

= variable welle in the $$p_mig files up to wave bf, syear in bgp_mig

jobpr Job agreement at entry

Generated Variable. Source variables see Table 5 to Table 7.

1 Yes (undiff.: only stays outside Germany, and stays in Germany before bf)
2 Prospective job (since bf)
3 Employment contract (since bf)
4 Job as self-employed (since bf)
5 No (since be)
6 Did not look for job (since bf)
7 Does not apply, was a child (since bf)
-5 If question was not included in wave or in main filter path

lastmig starty of last move

= -2 for German-born respondents

lastmig_imp starty_imp of last move

= -2 for German-born respondents

lfgroup Labor force group

Generated Variable. Source variables see Table 5 to Table 7.

1 Seasonal worker, contract for work and labor
2 Highly qualified and experts with special entry conditions
3 Qualified labor force with priority check by the Fed. Work Agency (not: bd, be)
4 Other labor force with priority check by the Fed. Work Agency (not: bd, be)
5 Trainee, au pair (not bd)
6 Self-employed, entrepreneur
7 Other
8 Relocated to Germany by employer (only bd, be)
9 Sent to Germany by company (only bd, be)
-5 If question was not included in wave or in main filter path

migfilter Marker for German-born/abroad-born respondents

Generated variable, based on information on the birth country.

1 Born in Germany
2 Born abroad

move Type of move

= nmtype of preceding spell

nmtype Type of next move

Source variables see Table 5 and Table 6.

1 To Germany

2 Back to country of birth

3 To another country

nspells Number of spells of a person.

Generated variable

ostatus Status at entry ("other" open answers)

Generated from open answers to status-variables. Replaces variables ostatus, ostatus-de and ostatusoc in previous releases of MIGSPELL.

Open answers are captured for the status variables related to stays abroad. For stays in Germany, open answers were only captured in the first wave (bd).

In previous releases of MIGSPELL, coded open answers were assigned directly to the variable status (and statusde, statusoc), if they fitted a level.

From release v32 of MIGSPELL, ostatus is designed to mirror if a level fits to a level of status1 or status2, but no direct assignment to status1 or status2 has been made. The first digit of the two-digit codes in ostatus indicates if the level fits variables status1 or status2, the second digit indicates which level of status1 or status2 a certain level of ostatus fits. Example: open answers with the meaning "Spouse, child, or family member" were coded 23 in ostatus, because the standardized answers with the same meaning are assigned to code 3 of status2.

One-digit codes do not correspond to a level of status1 or status2

1 Visit of family or friends

2 Love attachment

3 Au pair, gap year spent on voluntary social work

4 Intern, trainee

5 Military service, soldier

11 German migrant from Eastern Europe

21 Labor force (not bd)

22 Labor force with job agreement at entry

23 Spouse, child, family member

24 Asylum seeker, refugee

25 Student, apprentice

26 Seeking for job

27 Tourist

28 With tourist visa

29 None of these / other

sample1 Sample Indicator

24 M1 2013 Migration (1995-2011)

29 M2 2015 Migration (2009-2013)

start Start date of a spell in months from jan 1900

Generated variable from starty_imp and startmo_imp

startmo Start month of an episode (original values)

Source variables see Table 5 and Table 6.

startmo_imp Start month of an episode (imputed values)

Missing values in startmo are replaced with imputed values, where possible (see section 4)

starty Start year of an episode. (original values)

Source variables see Table 5 and Table 6.

starty_imp Start year of an episode (imputed values)

Missing values in starty are replaced with imputed values, where possible (see section 4)

status1 Legal background of entry

Generated Variable. Mapping of source variables to levels of status1 see Table 7. p. 16.

status1 and status2 replace variables status, statusde, statusoc of former releases of MIGSPELL

1 German migrant from Eastern Europe
2 German citizen, grown-up outside Germany (since be)
3 EU-citizen (only be)
4 EU- or EEZ-citizen with right to free movement (since bf)
5 EU- or EEZ-citizen without right to free movement (since bf)
6 Other citizens
-5 If question was not included in wave or in main filter path

status2 Status at entry

Generated Variable. Mapping of source variables to levels of status2 see Table 7.

1 Labor force (not bd)
2 Labor force with job agreement at entry
3 Spouse, child, family member
4 Asylum seeker, refugee
5 Student, trainee
6 Seeking for job
7 Tourist
8 With tourist visa
9 None of these / other

staytime Duration of stay (months)

Generated from imputed start and end variables: = end - start + 1

stype Spell type (marker for type of stay)

Generated from variable move

1 Stay in Germany

2 Stay abroad

-2 In first spell (= episode from birth to first move)

tcountry Target country of the next move

Coding according to SOEP-convention

Source variables see Table 5 and Table 6.

2.5 The date-variables in MIGSPELL

With spell data, the information on start date and end date of each episode is crucial. The migration biographies constitute a special kind of spell data, without parallelities and gaps: Because the questionnaire does not allow for reporting more than one place of habitation at the same time, the spells of a person cannot be parallel to each other.

The participants were only asked at what time they moved to a certain country, but not at what time they left. As one necessarily has to stay at one place or another, the episodes are successive to each other, and the end of a spell could be derived from the start of the next spell. Therefore the migration biographies have no gaps. [7]

However, sometimes respondents were not able to recollect the exact dates of their moves, or they named dates that contradicted the time sequence of moves, which led to missing values in the date variables. The respondents were not forced to state a date, and time contradictions were not immediately clarified during the interview, so these kinds of missings were to be expected. Overall, in the actual release v33, 379 spells (=2.42 %) have a missing value in the original start year and/or start month[8].

In order to make those spells accessible for data analyses, missing values were replaced with imputed values to the extent possible, resulting in only 9 spells with missing values in the imputed version of the startyear and/or startmonth.

From the imputed versions of startyear and startmonth, the start and end variables were generated. These contain integers counting the months that have passed since January 1, 1900, (i.e. including January). This is in contrast to the SOEP standard, where the counting of time

[7] As episodes shorter than 3 months are not captured, possible gaps stemming from these get hidden by the described procedure.

[8] The update from v33 to v33.1 does not affect the content of MIGSPELL.

begins January 1, 1983, and also to the Stata standard, where counting begins January 1, 1960.

We changed the zero-point of the time scale in MIGSPELL because, otherwise, the SOEP-convention on missing codes (which are integers < 0 and > -10) would have conflicted: Some migration biographies begin earlier than January 1983 or 1960, which results in negative values if we had used the SOEP or Stata standard. In certain cases, missing and valid codes would then have become indistinguishable.

If you are using Stata for data analyses and want to benefit from the inbuilt Stata time and date functions, you can transform the `start` and `end` variables into the Stata standard by subtracting 271. Before doing so, you must replace the SOEP missing values with values that are distinguishable from valid values, such as the Stata missing codes:

```
gen start_stata = start
gen end_stata = end
recode start_stata end_stata (-1 = .a) (-3 = .b)
replace start_stata = start_stata - 271
replace end_stata = end_stata - 271
format start_stata end_stata %tmCCYY_mon
```

See section 4 for a detailed description of the imputation of missing data values in MIGSPELL and REFUGSPELL.

2.6 Structure tables: Sources of the MIGSPELL variables for each spell (actual wave)

Table 5: bgp MIGSPELL synopsis 1: Stays abroad (for German-born migrants)

Var##	country	starty	startmo	status1	status2	jobpr	lfgroup	nmtype	tcountry	coverage
01	bgpm_l_0303 -2 tnz: DE integr.	bgpm_l_0103	bgpm_l_0102	---			---	bgpm_l_34	bgpm_l_35	all cases

from here: repetition for each value of Var##

01		bgpm_l01_3601	bgpm_l01_3602		bgpm_l01_37	bgpm_l01_37	---	bgpm_l01_3801	bgpm_l01_3802	stay abroad
01		bgpm_l01_3901	bgpm_l01_3902		---	---	---	bgpm_l01_4001	bgpm_l01_4002	stay in DE
02		bgpm_l02_3601	bgpm_l02_3602		bgpm_l02_37	bgpm_l02_37	---	bgpm_l02_3801	bgpm_l02_3802	stay abroad
02		bgpm_l02_3901	bgpm_l02_3902		---	---	---	bgpm_l02_4001	bgpm_l02_4002	stay in DE

etc.

15		bgpm_l15_3601	bgpm_l15_3602		bgpm_l15_37	bgpm_l15_37	---	bgpm_l15_3801	bgpm_l15_3802	stay abroad
15		bgpm_l15_3901	bgpm_l15_3902		---	---	---	bgpm_l15_4001	bgpm_l15_4002	stay in DE (up to int.date)

Table 6: bgp MIGSPELL synopsis 2: Coming to Germany (for migrants not born in Germany)

Var##	country	starty	startmo	status1	status2	jobpr	lfgroup	nmtype	tcountry	coverage
01	bgpm_l_0303 -2 tnz: DE intgr.	bgpm_l_0103	bgpm_l_0102	---	---	---	---	bgpm_l01_0701	bgpm_l_0702	all cases
01		bgpm_l_0601	bgpm_l_0602	---	bgpm_l01_1601	bgpm_l01_1601	---	bgpm_l01_1701	bgpm_l01_1702	stay abroad (birth country → not DE)
01		bgpm_l_0601	bgpm_l_0602	bgpm_l_08	bgpm_l_09 bgpm_l_11	bgpm_l_13	bgpm_l_10 bgpm_l_12	bgpm_l01_1401	bgpm_l01_1402	stay in DE
01		bgpm_l01_1501	bgpm_l01_1502	---	bgpm_l01_1601	bgpm_l01_1601		bgpm_l01_1701	bgpm_l01_1702	stay abroad (birth country → DE → not DE)

from here: repetition for each value of Var##

02		bgpm_l01_1801	bgpm_l01_1802	bgpm_l01_19	bgpm_l01_20 bgpm_l01_22	bgpm_l01_24	bgpm_l01_21 bgpm_l01_23	bgpm_l02_1401	bgpm_l02_1402	stay in DE
02		bgpm_l02_1501	bgpm_l02_1502	---	bgpm_l02_1601	bgpm_l02_1601	---	bgpm_l02_1701	bgpm_l02_1702	stay abroad
03		bgpm_l02_1801	bgpm_l02_1802	bgpm_l02_19	bgpm_l02_20 bgpm_l02_22	bgpm_l02_24	bgpm_l02_21 bgpm_l02_23	bgpm_l03_1401	bgpm_l03_1402	stay in DE
03		bgpm_l03_1501	bgpm_l03_1502	---	bgpm_l03_1601	bgpm_l03_1601	---	bgpm_l03_1701	bgpm_l03_1702	stay abroad

etc.

15		bgpm_l14_1801	bgpm_l14_1802	bgpm_l14_19	bgpm_l14_20 bgpm_l14_22	bgpm_l14_24	bgpm_l14_21 bgpm_l14_23	bgpm_l15_1401	bgpm_l15_1402	stay in DE
15		bgpm_l15_1501	bgpm_l15_1502	---	bgpm_l15_1601	bgpm_l15_1601	---	bgpm_l15_1701	bgpm_l15_1702	stay abroad
16		bgpm_l15_1801	bgpm_l15_1802	bgpm_l15_19	bgpm_l15_20 bgpm_l15_22	bgpm_l15_24	bgpm_l15_21 bgpm_l15_23	---	---	stay in DE (up to int.date)

2.7 Concordance of source and MIGSPELL variables by their values (wave bf, bg)

The following Table 7 shows a minute synopsis of the source- and target variables level by level for waves bf and bg. It overlaps with Table 16 in section 5, which shows the same for waves bd to bf. Other than from waves bd to bf, from wave bf to wave bg no structural changes occurred. At first glance, the variable names seem identical except for the wave indicator. However, this is only true for most of the variables, not for all of them.

Table 7: Detailed mapping of the source variables to the MIGSPELL variables for waves bf and bg

Note on the representation of the variables in Table 7:

The numerals (with four-digit numerals: the first two digits) after the last low line in the variable names indicate the number of the related question in the wave-specific questionnaire. The cipher signs (##) are wild-cards for the loop numerator in the variable name.

The numbers in parenthesis indicate the level of the source variables that enter in the respective level of the MIGSPELL variables.

The different colors of the variable names refer to different filter paths for different respondent groups and types of moves: **black**: Born outside Germany and a move to Germany, **blue**: Born outside Germany and a move abroad, **red**: Born in Germany and a move abroad.

bfp_mig	**bgp_mig**	**MIGSPELL**	**signification**
status1			
bfpm_l_08 (-1) bfpm_l##_19 (-1)	bgpm_l_08 (-1) bgpm_l##_19 (-1)	status1 : -1	
bfpm_l_08 (1) bfpm_l##_19 (1)	bgpm_l_08 (1) bgpm_l##_19 (1)	status1 : 1	German migrant from Eastern Europe
bfpm_l_08 (2) bfpm_l##_19 (2)	bgpm_l_08 (2) bgpm_l##_19 (2)	status1 : 2	German citizen, grown-up outside Germany
		status1 : 3	EU-citizen
bfpm_l_08 (3) bfpm_l##_19 (3)	bgpm_l_08 (3) bgpm_l##_19 (3)	status1 : 4	EU- or EEZ-citizen with right to free movement
bfpm_l_08 (4) bfpm_l##_19 (4)	bgpm_l_08 (4) bgpm_l##_19 (4)	status1 : 5	EU- or EEZ-citizen without right to free movement
bfpm_l_08 (5) bfpm_l##_19 (5)	bgpm_l_08 (5) bgpm_l##_19 (5)	status1 : 6	Other citizens

status2			
bfpm_l_09 (1) bfpm_l_11 (1) bfpm_l##_20 (1) bfpm_l##_22 (1)	bgpm_l_09 (1) bgpm_l_11 (1) bgpm_l##_20 (1) bgpm_l##_22 (1)	status2 : 1	Labor force
bfpm_l##_1601 (1) bfpm_l##_3701 (1)	bgpm_l##_1601 (1) bgpm_l##_37 (1)	status2 : 2	Labor force with job agreement at entry
bfpm_l_09 (4) bfpm_l_11 (4) bfpm_l##_20 (4) bfpm_l##_22 (4) bfpm_l##_1601 (2) bfpm_l##_3701 (2)	bgpm_l_09 (4) bgpm_l_11 (4) bgpm_l##_20 (4) bgpm_l##_22 (4) bgpm_l##_1601 (2) bgpm_l##_37 (2)	status2 : 3	Spouse, child, family member
bfpm_l_11 (5) bfpm_l##_22 (5) bfpm_l##_1601 (3)	bgpm_l_11 (5) bgpm_l##_22 (5) bgpm_l##_1601 (3)	status2 : 4	Asylum seeker, refugee
bfpm_l_09 (3) bfpm_l_11 (3) bfpm_l##_20 (3) bfpm_l##_22 (3) bfpm_l##_1601 (4) bfpm_l##_3701 (3))	bgpm_l_09 (3) bgpm_l_11 (3) bgpm_l##_20 (3) bgpm_l##_22 (3) bgpm_l##_1601 (4) bgpm_l##_37 (3))	status2 : 5	Student, trainee
bfpm_l_09 (2) bfpm_l_11 (2) bfpm_l##_20 (2) bfpm_l##_22 (2) bfpm_l##_1601 (5) bfpm_l##_3701 (4)	bgpm_l_09 (2) bgpm_l_11 (2) bgpm_l##_20 (2) bgpm_l##_22 (2) bgpm_l##_1601 (5) bgpm_l##_37 (4)	status2 : 6	Seeking for job
bfpm_l_09 (5) bfpm_l_11 (6) bfpm_l##_20 (5) bfpm_l##_22 (6)	bgpm_l_09 (5) bgpm_l_11 (6) bgpm_l##_20 (5) bgpm_l##_22 (6)	status2 : 7	Tourist
		status2 : 8	with tourist visum ***only in wave be***
bfpm_l_09 (6) bfpm_l_11 (7) bfpm_l##_20 (6) bfpm_l##_22 (7) bfpm_l##_1601 (6) bfpm_l##_3701 (5)	bgpm_l_09 (6) bgpm_l_11 (7) bgpm_l##_20 (6) bgpm_l##_22 (7) bgpm_l##_1601 (6) bgpm_l##_37 (5)	status2 : 9	None of these / other

jobpr			
bfpm_l##_1601 (1) bfpm_l##_3701 (1)	bgpm_l##_1601 (1) bgpm_l##_37 (1)	jobpr : 1	Yes (undiff.)
bfpm_l_13 (1) bfpm_l##_24 (1)	bgpm_l_13 (1) bgpm_l##_24 (1)	jobpr : 2	Prospective job
bfpm_l_13 (2) bfpm_l##_24 (2)	bgpm_l_13 (2) bgpm_l##_24 (2)	jobpr : 3	Employment contract
bfpm_l_13 (3) bfpm_l##_24 (3)	bgpm_l_13 (3) bgpm_l##_24 (3)	jobpr : 4	Job as self-employed
bfpm_l_13 (4) bfpm_l##_24 (4)	bgpm_l_13 (4) bgpm_l##_24 (4)	jobpr : 5	No
bfpm_l_13 (5) bfpm_l##_24 (5)	bgpm_l_13 (5) bgpm_l##_24 (5)	jobpr : 6	Did not look for job
bfpm_l_13 (6) bfpm_l##_24 (6)	bgpm_l_13 (6) bgpm_l##_24 (6)	jobpr : 7	Does not apply, was a child
lfgroup			
bfpm_l_10 (1) bfpm_l_12 (1) bfpm_l##_21 (1) bfpm_l##_23 (1)	bgpm_l_10 (1) bgpm_l_12 (1) bgpm_l##_21 (1) bgpm_l##_23 (1)	lfgroup : 1	Seasonal worker, contract for work and labor
bfpm_l_12 (2) bfpm_l##_23 (2)	bgpm_l_12 (2) bgpm_l##_23 (2)	lfgroup : 2	Highly qualified and experts with special entry conditions
bfpm_l_12 (3) bfpm_l##_23 (3)	bgpm_l_12 (3) bgpm_l##_23 (3)	lfgroup : 3	Qualified labor force with priority check by the Fed. Work Agency
bfpm_l_12 (4) bfpm_l##_23 (4)	bgpm_l_12 (4) bgpm_l##_23 (4)	lfgroup : 4	Other labor force with priority check by the Fed. Work Agency
bfpm_l_10 (3) bfpm_l_12 (5) bfpm_l##_21 (3) bfpm_l##_23 (5)	bgpm_l_10 (3) bgpm_l_12 (5) bgpm_l##_21 (3) bgpm_l##_23 (5)	lfgroup : 5	Trainee, au pair
bfpm_l_10 (4) bfpm_l_12 (6) bfpm_l##_21 (4) bfpm_l##_23 (6)	bgpm_l_10 (4) bgpm_l_12 (6) bgpm_l##_21 (4) bgpm_l##_23 (6)	lfgroup : 6	Self-employed, entrepreneur
bfpm_l_10 (5) bfpm_l_12 (7) bfpm_l##_21 (5) bfpm_l##_23 (7)	bgpm_l_10 (5) bgpm_l_12 (7) bgpm_l##_21 (5) bgpm_l##_23 (7)	lfgroup : 7	Other
		lfgroup : 8	Relocated to Germany by employer ***only in waves bd and be***
		lfgroup : 9	Sent to Germany by company ***only in waves bd and be***

3 REFUGSPELL v33

3.1 Main Differences between MIGSPELL and REFUGSPELL

- In MIGSPELL, the rule to report only stays with a minimum duration of three months is observed with very few exceptions. In REFUGSPELL, this is not the case. Many respondents reported shorter stays in transition countries, despite interviewer instructions to capture only stays that were at least three months long. However, we do not know to what extent the respondents reporting only long-term stays could have reported short-term stays if they had not observed the instruction. Nevertheless, we did not remove the short term stays from the migration biographies; instead including a flag variable to indicate very short stays of less than a month.
- The Refugees Survey posed some challenges with regard to language problems. Some respondents were interviewed with the help of a CAPI in their native language and lettering. Further, it seems in some cases that the respondent took over the laptop and filled in the answers by himself, without intervention of the interviewer. This shows in the data, where, in contrast to MIGSPELL, quite a few migration biographies are reported in an obviously wrong time sequence or with the same episodes reported twice. We cleaned the biographies if such error types could be determined clearly and included a flag variable to indicate if a reported migration biography was adjusted while generating REFUGSPELL.
- While MIGSPELL focuses differentiatedly on the terms under which a person *entered* a country (variables status1, status2, ostatus, jobpr, lfgroup), REFUGSPELL only specifies if a person *left* a country other than Germany for escape or for other reasons, without further particulars.
- In MIGSPELL and in REFUGSPELL, the last move to Germany is derived from the migration biography. However, other than the migration questionnaire, the refugees' questionnaire captures the date of the last move to Germany a second time in an extra question outside of the migration biography. Predictably, this leads to conflicting information on the date of the last move to Germany in some cases. As REFUGSPELL is about the migration biography, we did not correct the date of the last move with information from that extra question, except if this date was missing in the migration biography.
- MIGSPELL contains the migration biographies of German-born and foreign-born migrants. In contrast, REFUGSPELL is only about foreign-born refugees. However, there are a few respondents (N=11 out of 4,816) who seemed to have reported being born in Germany. Of these, 6 respondents reported a first move from their birth country - seemingly Germany - to Germany. We did not correct for this, but set a flag in the variable f_bcountry_iffy.

3.2 Summary description of REFUGSPELL

REFUGSPELL is derived from the migration biographies, which are collected from each new respondent of the IAB-BAMF-SOEP Refugees Survey, samples M3 and M4.[9] The reported moves of the respondents were captured through a loop structure of the questionnaire, with the number of loops limited to 15. In comparison to MIGSPELL, the loop structure of the migration biographies in bgp_refugees is much less complex.

REFUGSPELL, v33.1, contains data of 4,527 respondents and a total of 12,049 spells[10]. On average, 1.66 moves were reported by each respondent in REFUGSPELL, compared to 1.41 moves by each foreign-born respondent in MIGSPELL.

3.3 Synopsis of the variables in REFUGSPELL (systematic order)

Variable	Variable label	Description
Identifiers		
hhnr	Original Household Number	Identifier throughout the SOEP
persnr	Never Changing Person ID	Identifier throughout the SOEP
mignr	Running no. of move	REFUGPELL-specific identifier. The combination of persnr and mignr identifies a spell. Runs from 0 because first spell is not a move but episode from birth to first move or interview date, respectively.
sample1	Sample Indicator	Sample indicator throughout the SOEP
Time-invariant Characteristics		
intyear	Interview year	Year when migration biography was surveyed
migfilter	German-born/abroad-born	in REFUGSPELL: only abroad-born (see note (1) on page 22)
bcountry	Country of birth	
lastmig	Year of last move to Germany	= starty of the last spell of a person
lastmig_imp	Year of last move to Germany (imputed version)	= starty_imp of the last spell of a person (see note (2) on page 22)
lastmig_refug	Year of last move to Germany (extra question)	Information used from Question 34, variable bgpr_l_3401 of bgp_refugspell

[9] See http://www.diw.de/en/diw_01.c.538695.en/research_advice/iab_bamf_soep_survey_of_refugees_in_germany.html for a description of the IAB-BAMF-SOEP Refugees Survey.
[10] The first release of refugspell in the SOEP V33 distribution contained data of 4,816 persons. 289 interviews turned out to be irregular, and were eliminated from V33.1, see https://www.diw.de/sixcms/detail.php?id=diw_01.c.574033.en

Time-variant Characteristics		
starty	Start year	Start year of an episode. Birth year in first spell of a respondent.
startmo	Start month	Start month of an episode. Birth month in first spell of a respondent.
starty_imp	Start year imputed	Same as starty, with imputed missing values and birth year updated with birth year from PPFAD
startmo_imp	Start month imputed	Same as startmo, with imputed missing values and birth month updated with birth month from PPFAD
start	Start (months from jan 1900)	generated from starty_imp and startmo_imp
end	End (months from jan 1900)	= start of next spell - 1. See note (3) on page 22
stype	Spell type	Move to Germany / move abroad
country	Country of the stay	
rexit	Reasons for emigration	source variables see Table 8 and Table 9.
move	Type of move	to Germany / back to birth country / to another country
nmtype	Type of next move	to Germany / back to birth country / to another country
tcountry	Target country of next move	
staytime	Duration of stay (months)	end - start + 1 (generated from imputed date variables)
Technical Variables (time variant and time invariant)		
censor	Censor	Censor Variable generated from original dates
censor_imp	Censor imput. version	Censor Variable generated from imputed dates
nspells	Number of spells	Number of moves of a respondent + 1 (= mignr(max) + 1)
f_birth-date_corr	Flag: birth date updated from PPFAD	= 1 if starty_imp and/or startmo_imp of first spell of a person is updated with PPFAD if a later interview gave different birth-date information
f_bio_corr	Flag: corrections in migbio	= 1 for all spells of a respondent, If one or more spells have been removed from the migration biography.
f_starty_imp	Flag: starty imputed	= 1 if missing starty has been imputed
f_startmo_imp	Flag: startmo imputed	= 1 if missing startmo has been imputed
f_bcountry_iffy	Flag: info birth country dubious	see note (1) below

f_short	Flag: more than one spell begin in same month	= 1 if the next spell begins in the same month

Notes:

(1) The target population of samples M3 and M4 are foreign-born persons, and being foreign-born was a screening criterion for being surveyed using the refugees questionnaire. Actually, very few respondents (11 of 4,527) reported being born in Germany. Of these, 6 reported their first move going from their birth country to Germany, which is an obvious contradiction. In these 6 cases, the flag variable f_bcountry_iffy is set to 1. For the other 5 respondents, we assume that the information on their birth country is true. At least it cannot be determined otherwise. Nevertheless variable migfilter is set to 1 "abroad-born" for all cases in REFUGSPELL, because there were no different filter paths for foreign/german-borns, as there was in MIGSPELL.

(2) If there were missing values in startyear and/or startmonth of the last spell, the variables bgpr_l_3401 and/or bgpr_l_3402 (Question: When did you arrive in Germany?) were used to replace the missing value for variable lastmig_imp.

(3) In general, end is derived from the start month of the next spell by subtracting 1 (→ no overlapping of spells). But in REFUGSPELL we find quite a few short spells, with the next spell beginning in the same month as the current spell. In those cases, the usual algorithm would have led to an end month that is earlier than the start month. In these cases, the end of a spell was generated as the start of the next spell, instead of: start of the next spell minus 1. Figuratively, the start and end months of short spells had to be squeezed into the same month as there was no "place" to apply the normal algorithm. As an indicator for this peculiarity, the flag variable f_short is set to 1.

3.4 Levels and value labels of the variables in REFUGSPELL (alphabetical order)

bcountry Birth country

Coding according to SOEP-convention.

Source variables see Table 5 and Table 6.

censor Censor Marker

Generated variable, based on the original start variables

0 Uncensored
1 Right censored
2 Right miss. censored
3 Left censored
4 Left and right censored
5 Left and right miss. censored
6 Left miss. censored
7 Left miss. and right censored
8 Left miss. and right miss. censored

censor_imp Censor Marker, imputed version.

Generated variable, based on the imputed start variables. Same levels as censor

country Country of the stay.

Coding according to SOEP-convention, = tcountry of preceding spell

end End date of a spell as months from jan 1900

Generated variable = start of the next spell - 1

In last spell, end is generated from interview date (welle in $$p_mig for waves bd - bf, syear in wave bg, and bdpm_pmonin, $$pm_monin for waves be and bf, bgpmmonin)

For short spells (= next spell begins in the same month than the actual spell) see note (3) on page 22.

f_bcountry_iffy Flag variable

see note (1) on page 23

f_birth-date_corr Flag variable

= 1 if starty and/or startmo of first spell was updated with birth-date from PPFAD

f_bio_corr Flag variable

= 1 for all spells of a respondent if corrections of his/her migration biography affected the original number of spells, in other words, if a duplicate spell was deleted.

f_short Flag variable

= 1 if the next spell begins in the same month as the actual spell.

f_startmo_imp Flag variable

= 1 if missing value in startmo was replaced by imputed value

f_starty_imp Flag variable

= 1 if missing value in starty was replaced by imputed value

intyear Year when the migration biography was surveyed

= variable welle in the $$p_mig files up to wave bf, syear in bgp_mig

lastmig starty of last move before inteview

Generated from the original starty

lastmig_imp starty_imp of last move before inteview

Generated from the imputed starty_imp

lastmig_refug Generated from Question 34 "When did you arrive in Germany"

= value of bgpr_l_3401 from bgp_mig

migfilter Marker for German-born/abroad-born respondents

Dichotomous variable, from information on the birth country. See note(1) on page 22.

2 Born abroad

move Type of move

= nmtype of preceding spell

nmtype Type of next move

Source variables see Table 5 and Table 6.

1 To Germany

2 Back to country of birth

3 To another country

nspells Number of spells of a person.

Generated variable

rexit reason why country was left

Generated Variable. Mapping of source variables to levels of rexit see Table 7.

1 Escape

2 Other reasons

sample1 Sample Indicator

24 M1 2013 Migration (1995-2011)

29 M2 2015 Migration (2009-2013)

start Start date of a spell in months from jan 1900

Generated variable from starty_imp and startmo_imp

startmo Start month of an episode (original values)

Source variables see Table 5 and Table 6, page 16.

startmo_imp Start month of an episode (imputed values)

Missing values in startmo are replaced with imputed values, where possible (see section 4)

starty Start year of an episode. (original values)

Source variables see Table 5 and Table 6, page 16.

starty_imp Start year of an episode (imputed values)

Missing values in starty are replaced with imputed values, where possible (see section 4)

staytime Duration of stay (months)

Generated from imputed start and end variables: = end - start + 1

stype Spell type (marker for type of stay)

Generated from variable move

1 Stay in Germany

2 Stay abroad

-2 In first spell (= episode from birth to first move)

tcountry Target country of the next move

Coding according to SOEP-convention

Source variables see Table 5 and Table 6, page 16.

3.5 The date variables in REFUGSPELL

The description of the date variables in MIGSPELL given in section 2.5 applies fully to the date variables in REFUGSPELL. Against that background, some differences must be explained.

Namely, in REFUGSPELL, about 2.6 % of the spells have a duration of less than a month, visible from the fact that the next spell begins in the same month than the actual spell. As explained in note (3) on page 22, the normal algorithm to generate the end of a spell was modified to accommodate these short spells. The flag variable f_short indicates spells of a duration of less than one month.

Also, other than in MIGSPELL, in REFUGSPELL the migration biographies of some respondents were chronologically disordered or had obviously duplicate spells. If only the chronological order was corrected, this was done with the imputed date variables (`starty_imp`, `startmo_imp`, `start`), so that the original values are still represented. If an obvious duplicate spell was removed, inevitably the original data was also removed. An indicator for this is a gap in the numbering of the spells in variable mignr, and the flag variable f_bio_corr set to 1 for the whole migration biography of that person.

3.6 Source variables of the REFUGSPELL variables (wave bg)

Table 8: bgp REFUGSPELL synopsis of the source variables for each spell

var#	country	starty	startmo	rexit	nmtype	tcountry	coverage
01	bgpr_l_0202	bgpr_l_0103	bgpr_l_0102	---	bgpr_l_06	bgpr_l01_1103	all cases
01		bgpr_l_0501	bgpr_l_0502	bgpr_l01_1201	bgpr_l01_1301	bgpr_l02_1103	stay abroad (birth country → not DE)
01		bgpr_l_0501	bgpr_l_0502		bgpr_l01_0801	bgpr_l01_1103	stay in DE
01		bgpr_l01_0901	bgpr_l01_0902	bgpr_l01_1201	bgpr_l01_1301	bgpr_l02_1103	stay abroad (birth country → DE → not DE)
from here: repetition for each value of Var##							
02		bgpr_l01_1501	bgpr_l01_1502		bgpr_l02_0801	bgpr_l02_1103	stay in DE
02		bgpr_l02_0901	bgpr_l02_0902	bgpr_l02_1201	bgpr_l02_1301	bgpr_l03_1103	stay abroad
03		bgpr_l02_1501	bgpr_l02_1502		bgpr_l03_0801	bgpr_l03_1103	stay in DE
03		bgpr_l03_0901	bgpr_l03_0902	bgpr_l03_1201	bgpr_l03_1301	bgpr_l04_1103	stay abroad
etc.							
15		bgpr_l14_1501	bgpr_l14_1502		bgpr_l15_0801	bgpr_l15_1103	stay in DE
15		bgpr_l15_0901	bgpr_l15_0902	bgpr_l15_1201	bgpr_l15_1301	bgpr_l16_1103	stay abroad
16		bgpr_l15_1501	bgpr_l15_1502		---	---	stay in DE (up to int.date)

Table 9: Source variables for the REFUGSPELL variable rexit

Within the loops that were designed to capture the migration biography, the respondents were asked in question 12 if they left an intermediate country (other than Germany) for escape or for other reasons. The reasons why the respondent left his/her birth country (Herkunftsland) were asked outside the loop, by question 36, in more detail.

In generating REFUGSPELL, the value of `rexit` in the first spell of each respondent was missing, because question 12 had not been asked for the very first move. The missing information had been replaced by the information from question 36 as displayed in the following table:

rexit		
bgp_refugees	REFUGSPELL	signification
	rexit: -2	Stays in Germany In the first spell of each respondent: Born in Germany
bdpm_l##_1201 (-1) In the first spell of each respondent: bgpr_l_3611 (-1)	rexit: -1	No answer to question 12 In the first spell of each respondent: No answer to question 36
bdpm_l##_1201 (1) In the first spell of each respondent: bgpr_l_3601 (1) and/or bgpr_l_3602 (1) and/or bgpr_l_3603 (1)	rexit: 1	Left country for escape reasons
bdpm_l##_1201 (2) In the first spell of each respondent: bgpr_l_3601 (-2) and bgpr_l_3602 (-2) and bgpr_l_3603 (-2) and bgpr_l_3611 (not equal -1)	rexit: 2	Left country for other reasons

4 The imputation of missing date values in the generation of MIGSPELL and REFUGSPELL

4.1 General remarks

As previously mentioned, the `start` and `end` variables are generated from the *imputed* startyear and startmonth information. The original variables `starty` and `startmo` are not touched, so that they still can be used for analyses instead of the imputed versions. All replacements of values were made in the variables `starty_imp` and `startmo_imp`.

Here we explain how the imputation was performed.

Before presenting the applied rules and algorithms for the replacement of missing dates, the treatment of the special case "birth-date" is described:

The birth-date is represented in `starty` and `startmo` of the first spell of a respondent, which relates to the episode from birth to the first move to another country. If missing values in the birth-date had existed (which was not the case), they would not have been subject to imputation. In the first spell, there is no floor for possible values; therefore, a span for the estimation is not determined. However, the birth-date was compared with the birth-date from PPFAD. The birth-date is surveyed in each wave anew and PPFAD holds the newest answer of the respondent. If there was a difference between the birth-date from `starty` and `startmo` with the one from PPFAD, `starty_imp` and `startmo_imp` was updated with the newer information, and the flag variables `f_startmo_imp` and `f_starty_imp`, respectively, were set to 1.

In one case, the difference was more than 28 years. As this is regarded as an improbable value change, the update of `starty_imp` and `startmo_imp` was undone.

4.2 Procedures for the imputation of the missing date values in MIGSPELL

In this section, the imputation of the startyear and startmonth for the missing dates is explained. The corresponding Stata syntax is too complex to be part of this documentation. In case of questions, please contact the author at kerhardt@diw.de.

In principle, the following description of the procedures applies to MIGSPELL as well as to REFUGSPELL. Due to certain differences between REFUGSPELL and MIGSPELL, the procedures for REFUGSPELL had to be modified in some aspects. These divergencies are explained in section 4.3.

For an overview of relevant differences between MIGSPELL and REFUGSPELL see section 3.1.

4.2.1 The imputation of missing values in the startyear

In a first step, the consistency of the startyears' sequence is tested and set to -3 if it is inconsistent; i.e. a startyear is earlier than the startyear of the previous spell. After that, all spells with a missing value (-1 or -3) are flagged. Then a marker is generated for spells with directly succeeding missing startyears.

For series of more than one spell with a missing startyear, rule 1 applies:

Rule 1: If there are more than one missing startyear in directly succeeding spells, they will not be imputed. Instead, the startmonth is also set to missing, namely, set to the value of the startyear (which is either -1 or -3).

Single spells with a missing value in startyear are treated as follows:

Rule 2: The missing startyear is replaced with the startyear of the next spell minus 1. If the next spell is the last spell of a case, the missing startyear is replaced by the interview year minus 1.

If the such imputed startyear is smaller than the startyear of the preceding spell, or if it is equal to the startyear of the preceding spell but the startmonth is smaller than the startmonth of the preceding spell (that is to say, if the imputed start time is earlier than the start time of the preceding spell), then the subtraction of 1 from the startyear of the next spell is undone, the imputed startyear equals the startyear of the next spell.

With this procedure, in a rare constellation, the spell with a missing startyear conflicts with the time sequence of either the preceding or the following spell. For example: The spell with a missing startyear has a startmonth "May," the preceding spell starts June, 2011, the following spell starts February, 2012. To handle this kind of constellation, a second test of the sequence consistency is performed, resulting in a reset of the originally missing, now imputed startyear to missing. In practice, mostly the startmonth is also missing if the startyear is missing, meaning that the chance that this occurs is very small.

Following rule 2, a spell with an imputed startyear lasts one year and 11 months maximally.

4.2.2 The imputation of missing values in the startmonth

The imputation of missing startmonths is only performed if the imputed startyear is not missing. Otherwise, there is no sense in keeping or imputing the startmonth information, and `startmo_imp` is set to the same missing code than `starty_imp` (either -1 or -3). Thus, the frame for the imputation of missing startmonths is always the startyear of the spell in question.

For the imputation of missing startmonths, rule 1 does not apply. More than one succeding missing startmonth may be imputed if the startyear is known, so that a lower and upper limit

for the missing data can be established. Multiple succeeding spells with missing startmonth information are referred to as a "series" of spells with a missing startmonth.

The principle of the imputation of a series of spells with missing startmonth is to portion the time between the last and the next established startmonth of the same year onto the spells with a missing startmonth that lay in between.

It has to be distinguished between cases where a) all spells in a year have missing startmonths and b) at least one spell in a year has a valid startmonth.

Considerations for case a) all spells of a year have missing startmonths:

Rule 3: 12 months are divided between the number of succeeding spells with a missing startmonth in the same year. The algorithms are:

a1) `startmo_imp = 1 + round(c * span)`

a2) `span = 12 / ctot + 1`

with:
c = running number of the spell with a missing startmonth (the succeeding spells with missing starmonth within a year numbered from 1 to n)
ctot = total of succeeding spells with missing startmonth within a year

EXAMPLE:
startmonth before imputation: -3, -1, -1
span = 12 / 3+1 = 3
startmonth after imputation: 4, 7, 10
(1 + 1 * 3 / 1 + 2*3 / 1 + 3*3)

NOTE: the algorithm means that only 11 spells with a missing startmonth can be placed within one year, although there is enough "place" for 12 spells, if each lasts 1 month. The reason for this is the "+ 1" in formula a1). Only by this, a single spell in a year with missing startmonth is assigned the desired result 7 instead of 6.

As the provisions for surveying the migration biographies say to capture only stays that lasted at least 3 months (which was not always met, however), the constellation of 12 spells with a missing startmonth in a year does not occur empirically.

Considerations for case b) there is at least one spell with a valid startmonth in the same year:

Rule 4 The adjacent spells with a valid startmonth in the same year form the lower and upper limits of the time span that can be divided between the number of succeeding spells with a missing startmonth in the same year. If a series of spells with a missing startmonth begin in the first or end in the last month in a year, the lower limit is 1 and the upper limit is 12, respectively. Otherwise, the lower limit is the

startmonth of the last spell before the series with missing startmonth, the upper limit is the startmonth of the next spell after the series with missing startmonth.

The algorithms are:

b1) `startmo_imp = lol + round(c * span)`

b2) `span = upl - lol + 1 / ctot + 1`

with:

lol = lower limit = startmonth of the last spell with valid startmonth before the series of spells with missing startmonth, or 1 respectively

upl = upper limit startmonth of the next spell with valid startmonth before the series of spells with missing startmonth, or 12 respectively

c and **ctot** as above.

We now understand, that formulas a1) and a2) are only special cases of formulas b1) and b2).

NOTE: The minimum permissible value for span is 1. In other words, between lower and upper limit there must be a gap of at least as many free time units (months) as there are spells with a missing startmonth in the series.

EXAMPLE:
startmonth before imputation: 2, -1, 3, 10
span = 10-2 / 2+1 = 8/3 = 2,66
imputed startmonths: 2 + round(1 * 2,66) = 5
2 + round(2 * 2,66) = 7
startmonth after imputation: 2, 5, 7, 10

After the imputation of missing values in `starty` and `startmo`, the variables `starty_imp` and `startmo_imp` are either both missing or both non-missing, because we decided a) to impute all missing startmonths if `starty` was valid and b) not to keep a nonmissing startmonth if `starty` was missing.

From the imputed variables `starty_imp` and `startmo_imp` the variable `start` is generated by:

```
gen start = cond(starty_imp > 0 & startmo_imp > 0, ///
      ((starty_imp-1900)* 12) + startmo_imp, starty_imp)
```

meaning: if `starty_imp` and `startmo_imp` are not missing, `start` is calculated as:

```
      starty_imp - 1900 * 12) + startmo_imp
```

otherwise it is equal to `starty_imp` (which is a missing code).

The **end** of a spell is generated by:

```
by persnr: gen end = cond(start[_n+1] > 0, start[_n+1]-1, start[_n+1])
```

meaning: end is calculated as the start of the next spell minus 1, provided that the start of the next spell is not missing and that the next spell belongs to the same case. In calculating end, the confinement "by persnr" is imperative, because the algorithm encroaches from the current onto the next spell.

4.3 On the adaptation of the imputation procedures to provide for different conditions in REFUGSPELL

One of the main differences between MIGSPELL and REFUGSPELL which impacted the imputation routines is the presence of very short spells in the latter data set. The procedures described above rely partly on gaps of more than one month between the start of a spell and the start of the next spell. As this was not given throughout all the spells of REFUGSPELL, the calculation of the available time space for the imputation of missing dates had to be adapted accordingly. This had a clear impact on the distribution of the imputed months. While in MIGSPELL a missing month of a single spell in a year is imputed as July, in REFUGSPELL it is imputed as June. As a result, the distribution of the imputed months in REFUGSPELL has a peak of 60% at value 6 (June), and in MIGSPELL a peak of 73% at value 7 (July).

A second deviation from the procedures for MIGSPELL concerns the date of the last move to Germany. In MIGSPELL the date of the eventual move to Germany is derived from the last move that was reported in the migration biography. In REFUGSPELL this is the same, but additionally there is data available from the question "When did you arrive in Germany?" (question 34).

If you have two questions in a survey that seem to ask for the same facts, discrepancies will occur inevitably. To maintain comparability to MIGSPELL, where not such an additional question was available, we refrained from systematically updating the information on the last move to Germany from the migration biography with the information of question 34. If, however, the date of the last migration to Germany was missing, we replaced it with the information from question 34 rather than imputing a value. So for the date of the last move to Germany the following priority list was employed:

bgpr_l##_1501 / bgpr_l##_1501 (start year / start month of last move to Germany)

if this was missing:

bgpr_l_3401 / bgpr_l_3402 (year / month from question "When did you arrive in Germany?"

if this was missing:

imputation of the year / month according to the algorithms that were developped for the imputation of missing data.

5 The integration of the migration biographies of the three waves 2013-2015 (bd, be, bf), for samples M1 and M2

The questionnaire for the survey of the migration biographies is different from wave to wave. The version for wave 2015 is expected to be a final version, but because legal regulations are involved, which may change in future, this is not set in stone. Regardless, the questionnaire for the new respondents of the M1 and M2 IAB-SOEP-Migration samples remained the same for the wave bg, whereas the questionnaire for the M3/M4 IAB-BAMF-SOEP-Migration sample (i.e. refugees) is very different, and it is neither sensible nor feasible to include the migration biographies of the M3/M4 sample into MIGSPELL.

An integration of different structures necessarily has to compromise. Consequently, some core variables of MIGSPELL - while mirroring the variables of the bdp_mig-file perfectly in the first wave - now have a more complex relationship to the $$p_mig files. For analyses and their correct interpretation, you should be aware of the information that goes - and does not go - into the different categories of the MIGSPELL variables. E.g., some categories systematically receive data only from a single wave, or from a certain group in wave bd, but from another group in wave bf.

To allow for keeping track of the relationship between the variables of the $$p_mig files and the MIGSPELL file, we provide the structure tables for German and abroad-born respondents and a synopsis of the levels of each variable of the migration biographies as a source for the core MIGSPELL variables for the past waves bd to bf in the following sections.

5.1 Structure Tables: $$p_mig-variables to MIGSPELL-variables for waves bd to bg

Table 10: bdp integrated MIGSPELL Synopsis 1: Stays abroad (for German-born migrants)

Var##	country	starty	startm	status1	status2	jobpr	lfgroup	nmtype	tcountry	coverage
01	bdpm_l_0203 -2 tnz: DE integr.	bdpm_l_0103	bdpm_l_0102	---			---	bdpm_l01_28	bdpm_l01_2902	all cases
from here: repetition for each value of Var##										
01		bdpm_l01_3001	bdpm_l01_3002	---	bdpm_l01_3101	bdpm_l01_3101	---	bdpm_l01_3201	bdpm_l01_3203	stay abroad
01		bdpm_l01_32a01	bdpm_l01_32a02	---	---	---	---	bdpm_l01_3301	bdpm_l01_3303	stay in DE
02		bdpm_l02_3001	bdpm_l02_3002	---	bdpm_l02_3101	bdpm_l02_3101	---	bdpm_l02_3201	bdpm_l02_3203	stay abroad
02		bdpm_l02_32a01	bdpm_l02_32a02	---	---	---	---	bdpm_l02_3301	bdpm_l02_3303	stay in DE
etc.										
15		bdpm_l15_3001	bdpm_l15_3002	---	bdpm_l15_3101	bdpm_l15_3101	---	bdpm_l15_3201	bdpm_l15_3203	stay abroad
15		bdpm_l15_32a01	bdpm_l15_32a02	---	---	---	---	bdpm_l15_3301	bdpm_l15_3303	stay in DE (up to int.date)

Table 11: bdp integrated MIGSPELL Synopsis 2: Coming to Germany (for migrants not born in Germany)

Var##	country	starty	startm	status1	status2	jobpr	lfgroup	nmtype	tcountry	coverage
01	bdpm_l_0203 -2 tnz: DE integr.	bdpm_l_0103	bdpm_l_0102	---			---	bdpm_l01_1701	bdpm_l01_1703	all cases
01		bdpm_l01_1601	bdpm_l01_1602	---	bdpm_l01_2201	bdpm_l01_2201	---	bdpm_l01_2301	bdpm_l01_2303	stay abroad (birth country → not DE)
01		bdpm_l01_1601	bdpm_l01_1602	bdpm_l01_1801	bdpm_l01_1801	bdpm_l01_1801	bdpm_l01_19	bdpm_l01_2001	bdpm_l01_2003	stay in DE
01		bdpm_l01_2101	bdpm_l01_2102	---	bdpm_l01_2201	bdpm_l01_2201	---	bdpm_l01_2301	bdpm_l01_2303	stay abroad (birth country → DE)

from here: repetition for each value of Var##

02		bdpm_l01_2401	bdpm_l01_2402	bdpm_l01_2501	bdpm_l01_2501	bdpm_l01_2501	bdpm_l01_26	bdpm_l02_2001	bdpm_l02_2003	stay in DE
02		bdpm_l02_2101	bdpm_l02_2102	---	bdpm_l02_2201	bdpm_l02_2201	---	bdpm_l02_2301	bdpm_l02_2303	stay abroad
03		bdpm_l02_2401	bdpm_l02_2402	bdpm_l02_2501	bdpm_l02_2501	bdpm_l02_2501	bdpm_l02_26	bdpm_l03_2001	bdpm_l03_2003	stay in DE
03		bdpm_l03_2101	bdpm_l03_2102	---	bdpm_l03_2201	bdpm_l03_2201	---	bdpm_l03_2301	bdpm_l03_2303	stay abroad

etc.

15		bdpm_l14_2401	bdpm_l14_2402	bdpm_l14_2501	bdpm_l14_2501	bdpm_l14_2501	bdpm_l14_26	bdpm_l15_2001	bdpm_l15_2003	stay in DE
15		bdpm_l15_2101	bdpm_l15_2102	---	bdpm_l15_2201	bdpm_l15_2201	---	bdpm_l15_2301	bdpm_l15_2303	stay abroad
16		bdpm_l15_2401	bdpm_l15_2402	bdpm_l15_2501	bdpm_l15_2501	bdpm_l15_2501	bdpm_l15_26	bdpm_l02_2001	bdpm_l02_2003	stay in DE (up to int.date)

Table 12: bep integrated MIGSPELL Synopsis 1: Stays abroad (for German-born migrants)

Var##	country	starty	startm	status1	status2	jobpr	lfgroup	nmtype	tcountry	coverage
01	bepm_l_0303 -2 tnz: DE integr.	bepm_l_0103	bepm_l_0102	---			---	bepm_l01_22	bepm_l01_2302	all cases

from here: repetition for each value of Var##

01		bepm_l01_2401	bepm_l01_2402		bepm_l01_2501	bepm_l01_2501	---	bepm_l01_2601	bepm_l01_2603	stay abroad
01		bepm_l01_2701	bepm_l01_2702		---	---	---	bepm_l01_2801	bepm_l01_2803	stay in DE
02		bepm_l02_2401	bepm_l02_2402		bepm_l02_2501	bepm_l02_2501	---	bepm_l02_2601	bepm_l02_2603	stay abroad
02		bepm_l02_2701	bepm_l02_2702		---	---	---	bepm_l02_2801	bepm_l02_2803	stay in DE

etc.

15		bepm_l15_2401	bepm_l15_2402		bepm_l15_2501	bepm_l15_2501	---	bepm_l15_2601	bepm_l15_2603	stay abroad
15		bepm_l15_2701	bepm_l15_2702		---	---	---	bepm_l15_2801	bepm_l15_2803	stay in DE (up to int.date)

Table 13: bep integrated MIGSPELL Synopsis 2: Coming to Germany (for migrants not born in Germany)

Var##	country	starty	startm	status1	status2	jobpr	lfgroup	nmtype	tcountry	coverage
01	bepm_l_0303 -2 tnz: DE intgr.	bepm_l_0103	bepm_l_0102	---	---	---	---	bepm_l01_0701	bepm_l01_0703	all cases
01		bepm_l_0601	bepm_l_0602	---	bepm_l01_1401	bepm_l01_1401	---	bepm_l01_1501	bepm_l01_1503	stay abroad (birth country → not DE)
01		bepm_l_0601	bepm_l_0602	bepm_l_08	bepm_l_08 bepm_l_0901 bepm_l_10 bepm_l01_11	bepm_l_0902	bepm_l01_11	bepm_l01_1201	bepm_l01_1203	stay in DE
01		bepm_l01_1301	bepm_l01_1302	---	bepm_l01_1401	bepm_l01_1401		bepm_l01_1501	bepm_l01_1503	stay abroad (birth country → DE → not DE)

from here: repetition for each value of Var##

02		bepm_l01_1601	bepm_l01_1602	bepm_l01_17	bepm_l01_17 bepm_l01_1801 bepm_l01_19 bepm_l01_20	bepm_l01_1802	bepm_l01_20	bepm_l02_1201	bepm_l02_1203	stay in DE
02		bepm_l02_1301	bepm_l02_1302	---	bepm_l02_1401	bepm_l02_1401	---	bepm_l02_1501	bepm_l02_1503	stay abroad

etc.

15		bepm_l14_1601	bepm_l14_1602	bepm_l14_17	bepm_l14_17 bepm_l14_1801 bepm_l14_19 bepm_l14_20	bepm_l14_1802	bepm_l14_20	bepm_l15_1201	bepm_l15_1203	stay in DE
15		bepm_l15_1301	bepm_l15_1302	---	bepm_l15_1401	bepm_l15_1401	---	bepm_l15_1501	bepm_l15_1503	stay abroad
16		bepm_l15_1601	bepm_l15_1602	bepm_l15_17	bepm_l15_17 bepm_l15_1801 bepm_l15_19 bepm_l15_20	bepm_l15_1802	bepm_l15_20	---	---	stay in DE (up to int.date)

Table 14: bfp integrated MIGSPELL Synopsis 1: Stays abroad (for German-born migrants)

Var##	country	starty	startm	status1	status2	jobpr	lfgroup	nmtype	tcountry	coverage
01	bfpm_l_0303 -2 tnz: DE integr.	bfpm_l_0103	bfpm_l_0102	---			---	bfpm_l01_34	bfpm_l01_3502	all cases

from here: repetition for each value of Var##

01		bfpm_l01_3601	bfpm_l01_3602		bfpm_l01_3701	bfpm_l01_3701	---	bfpm_l01_3801	bfpm_l01_3803	stay abroad
01		bfpm_l01_3901	bfpm_l01_3902		---	---	---	bfpm_l01_4001	bfpm_l01_4003	stay in DE
02		bfpm_l02_3601	bfpm_l02_3602		bfpm_l02_3701	bfpm_l02_3701	---	bfpm_l02_3801	bfpm_l02_3803	stay abroad
02		bfpm_l02_3901	bfpm_l02_3902		---	---	---	bfpm_l02_4001	bfpm_l02_4003	stay in DE

etc.

15		bfpm_l15_3601	bfpm_l15_3602		bfpm_l15_3701	bfpm_l15_3701	---	bfpm_l15_3801	bfpm_l15_3803	stay abroad
15		bfpm_l15_3901	bfpm_l15_3902		---	---	---	bfpm_l15_4001	bfpm_l15_4003	stay in DE (up to int.date)

Table 15: bfp integrated MIGSPELL Synopsis 2: Coming to Germany (for migrants not born in Germany)

Var##	country	starty	startm	status1	status2	jobpr	lfgroup	nmtype	tcountry	coverage
01	bfpm_l_0303 -2 tnz: DE intgr.	bfpm_l_0103	bfpm_l_0102	---	---	---	---	bfpm_l01_0701	bfpm_l01_0703	all cases
01		bfpm_l_0601	bfpm_l_0602	---	bfpm_l01_1601	bfpm_l01_1601	---	bfpm_l01_1701	bfpm_l01_1703	stay abroad (birth country → not DE)
01		bfpm_l_0601	bfpm_l_0602	bfpm_l_08	bfpm_l_09 bfpm_l_11	bfpm_l_13	bfpm_l_10 bfpm_l_12	bfpm_l01_1401	bfpm_l01_1403	stay in DE
01		bfpm_l01_1501	bfpm_l01_1502	---	bfpm_l01_1601	bfpm_l01_1601		bfpm_l01_1701	bfpm_l01_1703	stay abroad (birth country → DE → not DE)

from here: repetition for each value of Var##

02		bfpm_l01_1801	bepm_l01_1802	bfpm_l01_19	bfpm_l01_20 bfpm_l01_22	bfpm_l01_24	bfpm_l01_21 bfpm_l01_23	bfpm_l02_1401	bfpm_l02_1403	stay in DE
02		bfpm_l02_1501	bfpm_l02_1502	---	bfpm_l02_1601	bfpm_l02_1601	---	bfpm_l02_1701	bfpm_l02_1703	stay abroad
03		bfpm_l02_1801	bfpm_l02_1802	bfpm_l02_19	bfpm_l02_20 bfpm_l02_22	bfpm_l02_24	bfpm_l02_21 bfpm_l02_23	bfpm_l03_1401	bfpm_l03_1403	stay in DE
03		bfpm_l03_1501	bfpm_l03_1502	---	bfpm_l03_1601	bfpm_l03_1601	---	bfpm_l03_1701	bfpm_l03_1703	stay abroad

etc.

15		bfpm_l14_1801	bfpm_l14_1802	bfpm_l14_19	bfpm_l14_20 bfpm_l14_22	bfpm_l14_24	bfpm_l14_21 bfpm_l14_23	bfpm_l15_1401	bfpm_l15_1403	stay in DE
15		bfpm_l15_1501	bfpm_l15_1502	---	bfpm_l15_1601	bfpm_l15_1601	---	bfpm_l15_1701	bfpm_l15_1703	stay abroad
16		bfpm_l15_1801	bfpm_l15_1802	bfpm_l15_19	bfpm_l15_20 bfpm_l15_22	-bfpm_l15_24	bfpm_l15_21 bfpm_l15_23	---	---	stay in DE (up to int.date)

5.2 Summary description of the changes in MIGSPELL from v31 to v32

As some of the migration biography questions changed from wave to wave, it is necessary to map each wave separately to the target variables in MIGSPELL. As a result, the variables in the v32 release of MIGSPELL describing the conditions under which the migration took place have also changed.

This applies mainly to the former MIGSPELL variable `status`. It is now differentiated into `status1` and `status2`. In the first wave, the status when moving to Germany did not mirror all possible legal conditions of immigration. In the second wave, the question design changed to resolve this, but the adaptation of MIGSPELL was postponed until the current release, in view of further changes of the status-related questions in wave bf.

In addition, the former variables `statusde` and `statusoc`, which only differentiated status according to the type of the stay (a stay in Germany or a stay abroad), are abandoned in this release. If needed, these variables can easily be constructed by using the `stype` variable.

There are also changes in the open answers to the status-questions. If respondents choose the category "other," they were asked to specify. In wave bd, open answers were captured not only for moves abroad, but also for moves to Germany. In following waves, this was only the case for moves abroad. In this, v32, release of MIGSPELL the new variable `ostatus` replaces the former variables `ostatus`, `ostatusde` and `ostatusoc`, the two latter variables being redundant to `ostatus`, because, like `statusde` and `statusoc`, they only differentiated according to the type of the stay. Please note: A redesign of the construction and coding of `ostatus` has taken place (see section 2.4).

Variable `lfgroup` has received additional levels and the coding is changed.

A new variable `jobpr` is included. Further, several flag variables were added to indicate imputed date values.

The former variable `syear` is renamend `intyear`, for consistency reasons: In SOEP panel data and in SOEP long, `syear` is the wave indicator, running through years in which a respondent participated in the SOEP, or to which the data relates, respectively. In MIGSPELL, intyear is a constant for each respondent, because in the IAB-SOEP-migration samples, the migration biography is surveyed only once, usually in the first wave of participation.

Variable `support` is no longer part of MIGSPELL, because it is not a loop variable, but is related only to the last move to Germany of foreign-born respondents. If needed, this information can be merged to MIGSPELL from the $$p_mig data files (variables bdpm_l_27 for intyear 2013, bepm_l_21 for intyear 2014, and bfpm_l_2601, bfpm_l_2602, bfpm_l_2603 for intyear 2015).

5.3 Synopsis: Mapping of the migbiography variables of waves 2013-2015 to the MIGSPELL variables

Table 16, which follows, shows the mapping of the variables of the single waves bd through bf to the variables of MIGSPELL in full detail, level by level. Thus you can see at a glance, for instance, that level 3 of the MIGSPELL variable status1 receives data only from variables of wave be.

Annotation on the representation of the variables in Table 8:

- The numerals (in case of four-digit numerals: the first two digits of the numerals) after the last low line in the variable names indicate the numbers of the related question in the wave-specific questionnaire. The number signs (##) are wild-cards for the loop numerator in the variable name.
- The numbers in parenthesis indicate the levels of the source variables that entered in the respective levels of the MIGSPELL variables.
- The different colors of the variable names refer to different filter paths for different respondent groups and types of moves: black: Born outside Germany / a move to Germany, blue: Born outside Germany / a move abroad, red: Born in Germany / a move abroad.

Table 16: Synopsis of the migration biography variables as source variables for the MIGSPELL variables

bdp_mig	bep_mig	bfp_mig	MIGSPELL	signification
status1				
			status1 : -6	
bdpm_l01_1801 (all levels) bdpm_l##_2501 (all levels) bdpm_l01_2201 (all levels) bdpm_l##_3101 (all levels)	bepm_l_08 (4,6,7,8,9) bepm_l##_17 (4,6,7,8,9) bepm_l##_2501 (all levels)		status1 : -5	
			status1 : -3	
	bepm_l_08 (-2) bepm_l##_17 (-2)	bfpm_l_08 (-2) bfpm_l##_19 (-2) bfpm_l##_3701 (all levels)	status1 : -2	
	bepm_l_08 (-1) bepm_l##_17 (-1)	bfpm_l_08 (-1) bfpm_l##_19 (-1)	status1 : -1	
bdpm_l01_1801 (2) bdpm_l##_2501 (2)	bepm_l_08 (1) bepm_l##_17 (1)	bfpm_l_08 (1) bfpm_l##_19 (1)	status1 : 1	German migrant from Eastern Europe

bdp_mig	bep_mig	bfp_mig	MIGSPELL	signification
	bepm_l_08 (2) bepm_l##_17 (2)	bfpm_l_08 (2) bfpm_l##_19 (2)	status1 : 2	German citizen, grown-up outside Germany
	bepm_l_08 (3) bepm_l##_17 (3)		status1 : 3	EU-citizen
		bfpm_l_08 (3) bfpm_l##_19 (3)	status1 : 4	EU- or EEZ-citizen with right to free movement
		bfpm_l_08 (4) bfpm_l##_19 (4)	status1 : 5	EU- or EEZ-citizen without right to free movement
		bfpm_l_08 (5) bfpm_l##_19 (5)	status1 : 6	Other citizens
status2				
			status2 : -6	
			status2 : -5	
			status2 : -3	
			status2 : -2	
	bepm_l_08 (-1)		status2 : -1	
	bepm_l_0901 (2) bepm_l_10 (1) bepm_l##_1801 (2) bepm_l##_19 (1)	bfpm_l_09 (1) bfpm_l_11 (1) bfpm_l##_20 (1) bfpm_l##_22 (1)	status2 : 1	Labor force
bdpm_l01_1801 (1) bdpm_l##_2501 (1) bdpm_l##_2201 (1) bdpm_l##_3101 (1)	bepm_l##_1401 (1) bepm_l##_2501 (1)	bfpm_l##_1601 (1) bfpm_l##_3701 (1)	status2 : 2	Labor force with job agreement at entry
bdpm_l01_1801 (3) bdpm_l##_2501 (3) bdpm_l##_2201 (2) bdpm_l##_3101 (2)	bepm_l_08 (6) bepm_l_0901 (4) bepm_l##_17 (6) bepm_l##_1801 (4) bepm_l##_1401 (2) bepm_l##_2501 (2)	bfpm_l_09 (4) bfpm_l_11 (4) bfpm_l##_20 (4) bfpm_l##_22 (4) bfpm_l##_1601 (2) bfpm_l##_3701 (2)	status2 : 3	Spouse, child, family member
bdpm_l01_1801 (4) bdpm_l##_2501 (4) bdpm_l##_2201 (3)	bepm_l_08 (4) bepm_l##_17 (4) bepm_l##_1401 (3)	bfpm_l_11 (5) bfpm_l##_22 (5) bfpm_l##_1601 (3)	status2 : 4	Asylum seeker, refugee
bdpm_l01_1801 (5) bdpm_l##_2501 (5) bdpm_l##_2201 (4) bdpm_l##_3101 (3)	bepm_l_08 (7) bepm_l_0901 (3) bepm_l##_17 (7) bepm_l##_1801 (3) bepm_l##_1401 (4) bepm_l##_2501 (3)	bfpm_l_09 (3) bfpm_l_11 (3) bfpm_l##_20 (3) bfpm_l##_22 (3) bfpm_l##_1601 (4) bfpm_l##_3701 (3))	status2 : 5	Student, trainee

bdp_mig	bep_mig	bfp_mig	MIGSPELL	signification
bdpm_l01_1801 (6) bdpm_l##_2501 (6) bdpm_l##_2201 (5) bdpm_l##_3101 (4)	bepm_l_0901 (1) bepm_l01_11 (8) bepm_l##_1801 (1) bepm_l##_20 (8) bepm_l##_1401 (5) bepm_l##_2501 (4)	bfpm_l_09 (2) bfpm_l_11 (2) bfpm_l##_20 (2) bfpm_l##_22 (2) bfpm_l##_1601 (5) bfpm_l##_3701 (4)	status2 : 6	Seeking for job
		bfpm_l_09 (5) bfpm_l_11 (6) bfpm_l##_20 (5) bfpm_l##_22 (6)	status2 : 7	Tourist
	bepm_l_08 (8) bepm_l##_17 (8)		status2 : 8	with tourist visum
bdpm_l01_1801 (7) bdpm_l##_2501 (7) bdpm_l##_2201 (6) bdpm_l##_3101 (5)	bepm_l_08 (9) bepm_l_0901 (5) bepm_l##_17 (9) bepm_l##_1801 (5) bepm_l##_1401 (6) bepm_l##_2501 (5)	bfpm_l_09 (6) bfpm_l_11 (7) bfpm_l##_20 (6) bfpm_l##_22 (7) bfpm_l##_1601 (6) bfpm_l##_3701 (5)	status2 : 9	None of these / other
jobpr				
			jobpr : -6	
bdpm_l01_1801 (all levels except 1) bdpm_l##_2501 (all levels except 1) bdpm_l01_2201 (all levels except 1)	bepm_l_0902 (-2,-1,1,2) bepm_l##_1802 (-2,-1,1,2)		jobpr : -5	
			jobpr : -3	
		bfpm_l_13 (-2) bfpm_l##_24 (-2)	jobpr : -2	
		bfpm_l_13 (-1) bfpm_l##_24 (-1)	jobpr : -1	
bdpm_l01_1801 (1) bdpm_l##_2501 (1) bdpm_l01_2201 (1) bdpm_l##_3101 (1)	bepm_l_0902 (1) bepm_l##_1802 (1) bepm_l##_1401 (1) bepm_l##_2501 (1)	bfpm_l##_1601 (1) bfpm_l##_3701 (1)	jobpr : 1	Yes (undiff.)
		bfpm_l_13 (1) bfpm_l##_24 (1)	jobpr : 2	Prospective job
		bfpm_l_13 (2) bfpm_l##_24 (2)	jobpr : 3	Employment contract
		bfpm_l_13 (3) bfpm_l##_24 (3)	jobpr : 4	Job as self-employed
	bepm_l_0902 (2) bepm_l##_1802 (2)	bfpm_l_13 (4) bfpm_l##_24 (4)	jobpr : 5	No
		bfpm_l_13 (5) bfpm_l##_24 (5)	jobpr : 6	Did not look for job

bdp_mig	bep_mig	bfp_mig	MIGSPELL	signification
		bfpm_l_13 (6) bfpm_l##_24 (6)	jobpr : 7	Does not apply, was a child
lfgroup				
			lfgroup : -6	
			lfgroup : -5	
			lfgroup : -3	
			lfgroup : -2	
			lfgroup : -1	
bdpm_l01_19 (2) bdpm_l##_26 (2)	bepm_l01_11 (2) bepm_l##_20 (2)	bfpm_l_10 (1) bfpm_l_12 (1) bfpm_l##_21 (1) bfpm_l##_23 (1)	lfgroup : 1	Seasonal worker, contract for work and labor
bdpm_l01_19 (5) bdpm_l##_26 (5)	bepm_l01_11 (5) bepm_l##_20 (5)	bfpm_l_12 (2) bfpm_l##_23 (2)	lfgroup : 2	Highly qualified and experts with special entry conditions
		bfpm_l_12 (3) bfpm_l##_23 (3)	lfgroup : 3	Qualified labor force with priority check by the Fed. Work Agency
		bfpm_l_12 (4) bfpm_l##_23 (4)	lfgroup : 4	Other labor force with priority check by the Fed. Work Agency
	bepm_l01_11 (6,7) bepm_l##_20 (6,7)	bfpm_l_10 (3) bfpm_l_12 (5) bfpm_l##_21 (3) bfpm_l##_23 (5)	lfgroup : 5	Trainee, au pair
bdpm_l01_19 (1) bdpm_l##_26 (1)	bepm_l01_11 (1) bepm_l##_20 (1)	bfpm_l_10 (4) bfpm_l_12 (6) bfpm_l##_21 (4) bfpm_l##_23 (6)	lfgroup : 6	Self-employed, entrepreneur
bdpm_l01_19 (6) bdpm_l##_26 (6)	bepm_l01_11 (9) bepm_l##_20 (9)	bfpm_l_10 (5) bfpm_l_12 (7) bfpm_l##_21 (5) bfpm_l##_23 (7)	lfgroup : 7	Other
bdpm_l01_19 (3) bdpm_l##_26 (3)	bepm_l01_11 (3) bepm_l##_20 (3)		lfgroup : 8	Relocated to Germany by employer
bdpm_l01_19 (4) bdpm_l##_26 (4)	bepm_l01_11 (4) bepm_l##_20 (4)		lfgroup : 9	Sent to Germany by company