

SOEP Survey Papers

Series D – Variable Descriptions and Coding

SOEP – The German Socio-Economic Panel at DIW Berlin

2019

SOEP-Core v34 – The couple history files BIOCOUPLM and BIOCOUPLY, and marital history files BIOMARSM and BIOMARSY

Maik Hamjediers, Paul Schmelzer, Sascha-Christopher Geschke, and SOEP Group

Running since 1984, the German Socio-Economic Panel (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>

Editors:

Dr. Jan Goebel, DIW Berlin

Prof. Dr. Stefan Liebig, DIW Berlin and Freie Universität Berlin

Dr. David Richter, DIW Berlin

Prof. Dr. Carsten Schröder, DIW Berlin and Freie Universität Berlin

Prof. Dr. Jürgen Schupp, DIW Berlin and Freie Universität Berlin

Dr. Sabine Zinn, DIW Berlin

Please cite this paper as follows:

Maik Hamjediers, Paul Schmelzer, and Sascha-Christopher Geschke, and SOEP Group. 2019. SOEP-Core v34 – The couple history files BIOCOUPLM and BIOCOUPLY, and marital history files BIOMARSM and BIOMARSY. SOEP Survey Papers 742: Series D. Berlin: DIW/SOEP



This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

© 2019 by SOEP

ISSN: 2193-5580 (online)

DIW Berlin
German Socio-Economic Panel (SOEP)
Mohrenstr. 58
10117 Berlin
Germany

soeppapers@diw.de

SOEP-Core v34 – The couple history files BIOCOUPLM and BIOCOUPLY, and marital history files BIOMARSM and BIOMARSY

**Maik Hamjediers, Paul Schmelzer, and Sascha-Christopher Geschke, and SOEP
Group**

The couple history files **BIOCOUPLM** and **BIOCOUPLY**, and marital history files **BIOMARSM** and **BIOMARSY**

by Maik Hamjediers, Paul Schmelzer, and Sascha-Christopher Geschke

(This documentation is a new version of previous SOEP documentations for the same files and has benefited from the work made by previous generators. For readability reasons, we do not specifically cite and specify text that has been used directly from the older SOEP documents.)

With the BIOMARSM/Y and BIOCOUPLM/Y the SOEP provides consistent and continuous marital and partnership histories for nearly all adult respondents. Whereas BIOMARSM and BIOCOUPLM just build on the prospective information at the time of each interview, BIOMARSY and BIOCOUPLY – containing also retrospective data – provide complete marital histories of respondents, starting at the year of their birth. But since until wave 27 no questions on a respondents’ couple history were asked, BIOCOUPLY includes only those respondents who have answered the biography questionnaire in wave 28 or later (and those who have been observed since the age of 17). Thus, all four datasets contain different information and therefore Table 1 gives an overview on the most important differences.

Table 1: Overview of the datasets

	Provided type of history	Time unit	Starting time	Sample
BIOMARSM	Marital Histories	Month	Entry into SOEP	All adult SOEP-participants
BIOMARSY	Marital Histories	Annual	Year of birth	All adult SOEP-participants
BIOCOUPLM	Relationship Histories	Month	Entry into SOEP	All adult SOEP-participants
BIOCOUPLY	Relationship Histories	Annual	Year of birth	Adult respondents answered the biography questionnaire after wave 27 or which were observed since the age of 17

Note that the marital status in the \$PGEN data files, stored as \$FAMSTD, is derived from BIOCOUPLM and BIOMARSM at the time of the interview. Although the partner indicator PARTZ\$ supplied in the \$PGEN data files is considered in the generating process of BIOCOUPLM and BIOMARSM, due to different generating processes, it might not entirely match with \$FAMSTD. Furthermore, consistency checks between waves were done as well. That way, changes in \$FAMSTD between data distributions but also in the couple and marital history datasets are possible for former waves. Note also, that BIOCOUPLY(M) and BIOMARTS(Y/M) have “inclusive” endings, so the month of the ending still belongs to the

given spell. In contrast the “exclusive” endings imply that the month (if you use monthly spell data) of the given spell does not belong to the given spell.

Moreover, take into account that neither BIOCOUPL(Y/M) nor BIOMARS(Y/M) exist for the respective Refugee-Samples (M3, M4 and M5). No information on marital and partnership biographies have been collected for refugees to construct a parsimonious and reasonable survey instrument. Migration histories, for example, have been favored.

This documentation proceeds with a short description of the sources for the generating process and of the editing process of constructing logical and consistent marital and couple histories. These steps are important to understand the description of the four data files following in the next parts.

1 Sources of the couple and marital history

For the construction of individual marital and relationship histories we gathered information

1. on current marital *and couple* status conducted in the personal questionnaire \$P (and \$PAUSL and \$PLUECKE);
2. on monthly information on events, that may have occurred since the last personal interview, which are also stored in \$P (and \$PLUECKE);
3. on the generated partner pointer PARTZ\$\$ of the generated dataset \$PGEN, which links current partners living in the same household;
4. on the *marital and relationship biography* from the biographical questionnaire \$LELA (and BIOLELA).

The personal questionnaire comprises a question on the marital status at the month of interview (e.g. Question 147 in 2015), whereby information on registered partnerships is included only since 2011. Furthermore it is asked for the current couple status, which entails whether someone has a partner and if so whether they are living together (e.g. Question 148 and 149 in 2015). For immigrants we also used information on the marital status derived from the foreigner questionnaire stored until 1995 in \$PAUSL and for temporary drop outs we replaced missing information with data from a short version of the personal questionnaire, which information is stored in \$PLUECKE. Moreover at least once the majority of SOEP participants answers the biography instead of the personal questionnaire, and for those waves the current marital and couple status is stored in \$LELA. This hereby out of the four sources gathered information on the current marital couple status is used to generate spells, beginning at the time of the interview and reaching to the next interview in which a change in the status is reported.

For those interviewed again it is also asked for any changes that occurred during the last year (e.g. Question 173 in 2015).¹ This monthly information on the events ‘moved in together’, ‘marriage’, ‘divorce’, ‘separation’, ‘death of partner’ and since 2011 ‘starting a new relationship’ that may have occurred since the last personal interview supplements the previous generated marital and couple status spells. In addition, we added the generated partner pointer, which was just used as another source of. These three kind of information – events, the current status and the partner indicator – were used to generate the monthly datasets BIOMARSM and BIOCOUPLM.

For the annual files, we additionally resorted on the retrospective information of the biographical questionnaire, which nearly all respondents just answer one time while participating in the SOEP. Since until 2010 it was just asked for the last three marriages (e.g. Question 72 in 2010), for all respondents having used the old version it was only possible to generate marital and no couple biographies from their year of birth on. In 2011, design of the retrospective questionnaire collecting information on couples changed notably. It now contains information on up to three previous marriages, registered same-sex partnerships or long-term relationships, defined as lasting for at least six months (e.g. Question 83 in 2015). Hence, up to four relationships, are possible to mention.² Therefore for all respondents who have used the biography questionnaire after 2010 we generated the annual data BIOCOUPLY, whereas all responses to any biographical questionnaire were used to generate BIOMARSY. Naturally, the information drawn from the biographical questionnaires are preceded by the responses to the personal questionnaire for the following waves. This is done by extending the monthly spelldata from BIOCOUPLM and BIOMARSY to an annual format; thus for this time, also shorter relationships were possible to mention, for example via the monthly information on the events mentioned above.

2 Construction of marital histories

Information on marital history mainly stems from respondents’ retrospective reports on their own history. Thus, no other benchmark on the substance of these reports exists. This led to inconsistencies – for example, overlapping of reported marriages or impossible changes between legal states. In contrary to the relationship histories, where odd patterns of the reported histories occurred but were also possible for the most part, rules for logical histories of the legal marital status had to be laid down in the generation process:

¹ Due to the fact that events were collected retrospectively from first of January of the last calendar year until the month of interview, events in the beginning of a year could have been reported twice in two consecutive waves. Taking variations up to twelve months for the same event into account, the timing of the events were corrected.

² Please note, that also between 2011 and 2014 the questions of the biography questionnaire were changed. From 2011 to 2013 it was also possible to mention a fourth previous relationship, thus the questions for that relationship was limited compared to the question about the other relationships. Furthermore, the biography questionnaire of the National Survey of Families (FiD-Sample) left out the option of reg. same-sex partnerships.

1. Every individual marital history has to start with the state ‘unmarried’. We did not allow a person to be married before age 16.
2. From ‘unmarried’, one can only change to ‘married’ or into ‘living in a registered same-sex partnership’.
3. There is no possible return to ‘unmarried’ once a person was ever ‘married’.³ The only possible change from ‘married’ is to ‘divorced’, ‘widowed’ or ‘divorced or widowed’.
4. The only possible change from ‘divorced’ or ‘widowed’ is to ‘married’.
5. Reported separations of married spouses are taken into account as well. This led to the new category ‘married, separated’ which is coherent with the BIOCOUPL datasets and contains the time from a separation until a divorce or death of the respondent’s spouse. Yet, this rule was not applied for move-outs and later returns into the household of the same partner. If another marriage with a new partner is reported without a divorce or a death of the previous partner, after the ‘married, separated’-spell an episode of the category ‘divorced or widowed’ is also added and the beginning and end of those spells as well as the censoring were adjusted (Figure 1).

Figure 1: Example for treatment of separated marriages in BIOMARSM/Y

Respondents information						Added spells					
persnr	spellnr	spelltyp	begin	end	divorce	persnr	spellnr	spelltyp	begin	end	divorce
1	1	married	10	50	0	1	1	married	10	50	0
1	2	married	100	150	0	1	2	married, separated	50	-1	-1
						1	3	divorced or widowed	-1	100	0
						1	4	married	100	150	0

Algorithms rearrange and correct original data to produce logically consistent marital histories following these five rules. Thus, we inserted an obligatory first ‘unmarried’ spell starting with birth and ending with the first marriage. Possible contradictions in the data were checked and edited as well, i.e.: a) contradictions between responses given in the same year in the different, previously described sources; b) illogical sequences between years.

While most of the original data was used, we edited some spells, e.g. because of the notion of “romantic wedding”. This idea assumes that respondents often define their marital status in subjective or even affective terms rather than referring to the legal marital status. For instance, we frequently observe unmarried couples both reporting to be married for one year, and then returning to define themselves as singles (this phenomenon we call “romantic marriage”).

³ Except for the annulations of a marriage. However, given that this event is extremely rare – in particular compared to the many returns to ‘unmarried’ that we find in the data – we did not consider the possibility of annulations.

Another rather emotional defined marital status are changes between the reported state of being divorced and being married, but not living together. This might stand for an insecure period after a break-up of a marriage and the changes between the states are corrected until any clear indication for an actual divorce is reported. A third frequent pattern refers to divorced respondents who report to be unmarried after they started a new relationship, since they may want to avoid the new partner to come to know about the former marriage. Since this kind of change of the legal marital status is impossible, the status of being divorced is assigned till the next reported marriage.

If responses on marital status alternate between 'divorced' and 'widowed' without reporting another marriage in between, the information reported most often was used to correct these contradictions. Sequences alternating short-termed between 'divorced' / 'unmarried' and 'married' are replaced by 'married', as long as a new partner, a longer consistent sequence or the report of an event does not confirm this possibly new marriage. If the very end of a reported sequence indicates an 'unmarried' spell after being married and without further information, the marital history ends with a 'divorced or widowed', indicating that the reason for the end of a marriage is not known. Likewise, reported weddings during still existing marriages are ignored mostly.

In cases where information of both partners on their joint marriage is available, contradictions in dates were not dissolved and the original information is not replaced. You can identify these spells by sorting the couples in the BIOCPLM dataset using the variable COUPID and decide which information you find more reliable.

Completeness is a further criterion for construction of marital histories in the sense that the spell system is a closed system of spells starting from birth or the entrance into the SOEP going to the last year of sample membership. Due to item as well as partial unit non-response (i.e., a person of a SOEP households refuses to give a personal interview) and due to inconsistent information, 'gap' spells are introduced as another category of SPELLTYP on its own. Gap spells can occur at any place in the spell system, i.e., there are no restriction rules like the ones above. If information on marital status is missing for more than two years we inserted gap spells indicating that we have no knowledge of what happened during these periods. Likewise, missing retrospective biography information is indicated by an inserted gap spell. Also, if three finished marriages were reported in the biography questionnaire and a fourth current marriage via the next marital status, another gap is inserted. Missing information due to item non-response in the life course questionnaire may also affect the dates of the beginning or end of a spell.

3 Construction of couple histories

As stated above, most of reported odd patterns in the couple histories may be possible and hence, no verified decision between measurement error and uncommon reality can be made. For that reason, whenever possible, couple histories are left as they were stated. Only in rare cases, restrictions, corrections on orderings of events or changes of declared years are conducted (read the following paragraphs for details). Further corrections to smooth out irregularities are thus left to the user.

Firstly, the following default rules were obeyed to obtain logical and consistent histories if no other information forced to do otherwise:

1. Every individual couple history starts at the year of their birth with the state ‘single’. Because legal marriages are not possible before the age of 16, we restricted the age of marriage to be at least 16. Thus, we did not restrict age within a relationship, that is, unmarried relationships are allowed to start anytime.
2. Every spell set for a certain couple starts with the state ‘coupled, partner not in household’. One exception exists: if respondents report a year of moving together that lies before the start of their relationship, this specific couple history starts with ‘coupled, partner in household’. Note that in this case the information when this couple moved together is not available in BIOCPLM/Y anymore. Analogical the date of moving out is also lost if it was reported to come after the end of a relationship. Both dates you can still look it up in the original source.
3. If there is no evidence to the contrary, it is assumed that married couples live together and moved together before marriage. That is: for married couples their specific couple history starts with ‘coupled, partner not in household’ and is followed by a spell ‘coupled, partner in household’ before their marriage spell ‘married, spouse in household’ starts. Thus, if a couple moved together in the same year they married, a spell ‘coupled, partner in household’ is included anyway. The information whether spells needed to be inserted in order to this rule is stored in the variable REMARK (see Figure 2). Note that dates of becoming a couple and moving together or whether they moved together at all are not known in that case and were set equal to the time of marriage.

Figure 2: Example for additional spells before marriages

Respondents information						Added spells					
persnr	spellnr	spelltyp	begin	end	remark	persnr	spellnr	spelltyp	begin	end	remark
1	1	single	10	50	1	1	1	single	10	50	1
1	2	married, living together	50	100	1	1	2	coupled, not living together	50	50	2
						1	3	coupled, living together	50	50	2
						1	4	married, living together	50	100	1

- Since it was possible to mention another relationship in the questionnaire, it is assumed that periods between those relationships mentioned were 'single' states and thus assigned accordingly. Note that this relates to long-term relationships (longer than six months) reported via the biography questionnaire only. In contrast to those, relationships derived from reported changes in the familiar situation in the personal questionnaire may also be shorter than six month.
- Any (formerly) married couple that is not an active relationship anymore, i.e. married couple which is separated but not yet divorced, ends with a spell 'married-separated' (see Figure 3). As long as they are not divorced (or the partner hasn't died) yet, the end date of those separated spells is the same as their last interview year. Contrary to this, the separation spell is not added if marriage clearly ended with a divorce or the death of the respondent's partner and no previous break-up was reported. Be aware, that because the question for a separation does not differentiate between a move-out and a clear ending of a relationship, there is some uncertainty whether a marriage is still active, but not in a joint household or the relationship ended and the spouses are not divorced yet. By using the information of following interviews we identified the spelltyp as clear as possible. Note that 'married, separated' is a redundant spell in terms of couple history: It always and fully overlaps with other spells that contain the actual couple status(es) over the entire separation episode.

Figure 3: Example for treatment of separated marriages in BIOCOUPLM/Y

Respondents information						Added spells					
persnr	spellnr	spelltyp	begin	end	divorce	persnr	spellnr	spelltyp	begin	end	divorce
1	1	married, living together	10	50	0	1	1	married, living together	10	50	0
1	2	coupled, not living together	100	150	0	1	2	married, separated	50	150	-1
						1	3	single	50	100	0
						1	4	coupled, not living together	100	150	0

6. Because BIOCOUPLM/Y documents couple statuses and not marital statuses, it is possible to become single after marriage (in BIOMARSM/Y the only possible change from 'married is to 'divorced' or 'widowed').
7. Information on the partner's death or a divorce is stored in the variables PDEATH and DIVORCE. It is always just assigned to the last spell of a relationship; if e.g. a marriage ends in divorce, just the last spell containing also the definite end of the relationship is marked via DIVORCE and all previous spells for the same marriage (e.g. 'married, not living together' before a move-in) are marked as 'spell does not end with a divorce' (see also Figure 3). If not applicable, that is if a person is currently single, PDEATH and DIVORCE are set to '(-2) does not apply'. If a respective couple is not married, DIVORCE is set to '(-2), does not apply', as well. (Be careful to change missings into system missing values if you plan to sum up these indicators column-wise.)

Checks were applied for three kind of inconsistencies: a) information from the different sources (e.g. biography and personal questionnaire) for the same years may conflict, b) dates within a personal couple history might appear plausible or not and c) contradictions between information given by a respondent and her relationship to the head of household.⁴ In general it was tried to leave as much of the original information as it is given by the respondent as possible to avoid making assumption which users wouldn't follow and to allow for analyses of uncommon patterns as well.

Concerning the reported status(es) (a) and dates (b), it is possible that a person has multiple, overlapping relationships in a certain period. That is, relationships (in opposition to marriages) do not need to be consecutive. The variable SPELLNR provides the sorting order, whereby the most recent spell comes last, which represents the relationship that started most recently. If no clear dates were available, the sorting in SPELLNR reflects the order of reporting by the respondent. Thus, some overlaps and inconsistencies might still appear in the dataset, which you may want to remove.

Contradictions in both partners' answers (c) are left as they were, even if both were interviewed in the current wave. E.g. information on starting date or the date of marriage of their joint relationship often differs. For very few cases, there were even contradictions about the actual partnership between two individuals in the sample. E.g., if one part of a couple or the position to the head of household suggested a couple in a household, but one or both answered to be single or coupled with someone else (outside the household). Because it may depend on a research question which information is needed, we left irregularities within couples as reported by each spouse; neither date nor the relationship status is edited or

⁴ See \$PSTELL in the dataset \$PBRUTTP. This information was given by the head of household in the Household Questionnaire and processed in partner pointer PARTZ\$\$.

overwritten between both partners of a current couple. To check for contradictions you can link them via couple identifier COUPID or PARTNR\$\$.

Some respondents refused to answer some or all questions in the relationship part. To some extent, missing values and contradiction are similar in the generating sense. Hence, some of the following rules applied to missing values, may seem familiar:

1. If a respondent stated to be married or in a relationship, but it is not known whether the couple lived together as well, these relationship spells were set as ‘coupled, partner not living in the household’ and the respective REMARK was coded as ‘edited’.
2. Furthermore, if we had clear information about whether a respondent was married or not, but there was contradicting information about whether a respondent was in a relationship or not, we assigned these spells as ‘unknown’ (98). This applies also to the counterpart, that means the relationship status was definite, but we lacked information whether the couple was married or not.
3. In some cases it is possible that a respondent might have had earlier relationships, but they could not be named anymore due to the above mentioned restrictions of the biography questionnaire. Thus, it is not known whether there were other partners before the latest mentioned. In these cases, gaps were introduced, that is a spell ‘unit nonresponse’ (99) was inserted.
4. The relationship history of respondents, who haven’t filled out the biography questionnaire, were also supplemented with additional spells: from age 0 they are stated ‘single’; between this first single spell until the first mentioned relationship gaps are included as well.
5. For many non-interviewed persons’ information on the current relationship can be reconstructed, because the respective partner is available from his or her personal interview. Thus, those non-interviewed adults remain in the BIOCOUNPLM dataset, even though information is not directly retrieved from them. If information on their current couple status was given by their partner, it is fully copied to these non-respondents as well; the variable REMARK is set as ‘added’ for these spells.

4 BIOCOUNPLM: A monthly couple biography

The spell data in the file BIOCOUNPLM contains information on marital biographies starting with responses to the first personal interview. The data file comprises fourteen variables: the original household, couple and individual identifiers HHNR, COUPID and PERSNR as well as eleven specific variables. Please note for the couple identifier COUPID that the BEGIN and END the identified partner spells are not necessarily equal: Differences may occur because we give a preference to respondents’ own history. To achieve the same dates for

BEGIN and END one might give a preference to one of the partners (i.e. by using only the information reported by the women in the relationships).

Variable SPELLNR is a chronological index number for each individual's spells during the observation period. Variables BEGIN and END indicate the month in which a marital spell starts and ends. Monthly histories start with a value of 1 in January 1983 and ranges until the current margin in month 384, i.e., December 2014. In principle, the month in which a spell starts is the very same month the previous spell ends in. Please take into account that the months of BEGIN and END are not imputed when the exact date of the change in the familiar situation is not reported. Instead, the time of the interview is left as the BEGIN or END date when a change is observed in the data. We introduced therefore a variable called EVENTS, containing the information whether the exact month of the BEGIN or END of a spell is reported via the monthly information on events or not.

Variable SPELLTYP documents the partnership status with the possible categories listed below. Note that it is only asked for same-sex partnership since wave 28, thus, this information cannot be reconstructed for previous relationships. The state 'single' is not equal to the legal state 'never married'; it refers only to a person's current state of having a partner or not. 'Married, separated' spells refer only to couples who are still married, i.e. who are separated but not yet divorced. The spell 'Married, separated' always overlaps (in terms of starting and ending time) with other spells that contain the actual couple status(es) over the entire separation episode and is therefore redundant in terms of completeness of couple histories (e.g. spell "Coupled, partner in household" might overlap with spell 'Married, separated'). The additionally assigned codes 'implausible', 'unknown' and 'unit nonresponse' indicate implausible or a lack of information for the respective period.

Variables of BIOCOUPLM

HHNR	Identifier of original sample household
PERSNR	Personal identifier
COUPID	Couple identifier
SPELLNR	Consecutive spell number (chronological order)
SPELLTYP	Partnership status (-3) implausible (1) Married, spouse not in household (2) Married, spouse not in household (3) Coupled, partner in household (4) Coupled, partner not in household (5) Single (6) Married, separated (7) Registered same-sex partnership, partner in household (8) Registered same-sex partnership, partner not in household (98) Unknown (99) Unit nonresponse
BEGIN	Month when spell begins [1=Jan 1983 to 384=Dec 2014]
END	Month when spell ends [1=Jan 1983 to 384=Dec 2014]
BEGINY	Year spell begins [1983 to 2014; -3=implausible; -1=missing]
ENDY	Year spell ends [1983 to 2014; -3=implausible; -1=missing]
PDEATH	Death indicator: spell ends with death of partner?
DIVORCE	Divorce indicator: spell ends with divorce?
CENSOR	Censoring information [0 to 14] (see explanation below)
EVENTS	Month information: exact month of spell begin or end known?
REMARK	Further spell information (1) Original spell (2) Edited spell (3) Added spell (4) Gap Spell (5) First Spell

In addition, the indicator variables PDEATH and DIVORCE are provided. PDEATH indicates whether a respective spells ends with the death of a person's partner. Please note that this is not restricted to married persons, thus does not only refer to widowhood. DIVORCE is similar and indicates whether the marriage spell ends in divorce. Hence if it did not end in divorce, it is coded as missing when a new marriage has been reported or as a still ongoing marriage. In the second case END is updated by the month of the last interview. Both indicator variable PDEATH and DIVORCE are assigned with '(-2) does not apply' for single and '(-1) missing' for gap spells.

It is important to note that a 'first spell' in BIOCOUPLM is not the very first spell of a person, but the first observed partnership status since the person is taking part in the SOEP. Accordingly, the first spell in BIOCOUPLM is almost always left-censored, that means the beginning of the spell is not known. Variable CENSOR informs whether a spell is left- or right-censored (in fact most spells in BIOCOUPLM are censored). Table 2 gives an overview for the coding of the reasons responsible for censoring. With regards to left censoring we distinguished the reasons 'missing' (which also applies for most of the first spells), 'after gap'; for right censoring the reasons are categorized in 'missing', 'before gap', 'last spell',

and ‘death of the respondent’, whereas ‘death’ of course is always the ‘last spell’ and therefore the censor is overwritten with the former category. Certainly spells can also be coded as left- and right-censored combined.

Table 2: Coding of the variable CENSOR

Right: Left:	not censored	censored missing	censored before gap	censored last spell	censored death
not censored	0	3	4	5	6
censored missing	1	7	8	9	10
censored after gap	2	11	12	13	14

Note: ‘(99) Gap’, ‘(98) unknown’ and ‘(-3) implausible’ spells are all uncensored

Variable REMARK provides information on whether we had to edit or supplement original information provided by respondents in order to construct consistent couple biographies. Spells in BIOCOUNPLM are marked as ‘edited’ if the editing process involved substitution of reported information because of inconsistency with responses in previous or following interviews or information reported by a partner. Furthermore we have added spells, e.g. between two marriages because two marriages with separate persons at the same time are not legal (see former section for more details on editing); these spells are marked as ‘added spells’. Finally ‘first spells’ and ‘gap spells’ are marked separately, whereas also ‘added’ or ‘edited’ first spells are marked as ‘first spells’.

5 BIOCOUNPLY: An annual couple biography

The spells in the data file BIOCOUNPLY contain retrospectively and prospectively collected information on couple history since a respondent’s year of birth on an annual basis. Since until wave 27 no questions on a respondent’s couple history were asked retrospectively, BIOCOUNPLY includes only those respondents who have answered the biography questionnaire in wave 28. So except those, who were observed in the SOEP annually at least since the age of 17, people who answered the biography questionnaire before wave 28 are excluded from BIOCOUNPLY. Thus, BIOCOUNPLY comprises a much smaller sample than BIOCOUNPLM and the BIOMARSM/Y datasets. The newly developed retrospective part of the person questionnaire covers up to four relationships in addition to the current status.

The data files contain thirteen variables. In contrast to BIOCOUNPLM, BIOCOUNPLY just contains the original household and individual identifier, due to the fact that no information about the identity of each partner was surveyed in the biography questionnaire and therefore a partner cannot be identified for the relationships before the entrance into the SOEP. The following table provides an overview of the variables in BIOCOUNPLY, which contains beside the identifier eleven further variables.

Variables of BIOCOUPLY

HHNR	Identifier of original sample household
PERSNR	Personal identifier
SPELLNR	Consecutive spell number (chronological order)
SPELLTYP	Partnership status (-3) Implausible (1) Married, spouse in household (2) Married, spouse not in household (3) Coupled, partner in household (4) Coupled, partner not in household (5) Single (6) Married, separated (7) Registered same-sex partnership, partner in household (8) Registered same-sex partnership, partner not in household (98) Unknown (99) Unit nonresponse
BEGIN	Age of respondent when spell begins [-3=implausible; -1=missing]
END	Age of respondent when spell ends [-3=implausible; -1=missing]
BEGINY	Year when spell begins [-3=implausible; -1=missing]
ENDY	Year when spell ends [-3=implausible; -1=missing]
PDEATH	Death indicator: spell ends with death of partner?
DIVORCE	Divorce indicator: spell ends with divorce?
CENSOR	Censoring information [0 to 14] (see explanation above)
SOURCE	Source of information (1) derived only from biography questionnaire (2) derived from biography and personal questionnaire (3) derived only from personal questionnaire
REMARK	Further spell information (1) Original spell (2) Edited spell (3) Added spell (4) Gap Spell (5) First Spell

Variable SPELLTYP documents the partnership status with the possible categories listed above. Note that it was only asked for same-sex partnership since wave 28, thus, this information cannot retrospectively reconstructed for responds of the personal questionnaire. The state 'single' is not equal to the legal state 'never married'; it refers only to a person's current state of having a partner or not. 'Married, separated' spells apply only to former couples who are still married, i.e. who are separated but not yet divorced. It always overlaps with other spells that contain the actual couple status(es) over the entire separation episode and is therefore redundant in terms of completeness of couple histories. The additionally assigned codes 'implausible', 'unknown' and 'unit nonresponse' indicate implausibility or a lack of information for the respective period.

Variable SPELLNR is a chronological index number for each individual's spell during the observation period. Due to the fact that the spells' duration is measured in months via the personal questionnaire, it is important to note that an individual may encounter several events in the same year. In this case the variable SPELLNR allows the user to order spells with

respect to the respondent's life course. The variables `BEGIN` and `END` provide the years in which a spell begins and ends, whereas the variables `BEGIN` and `END` indicate respondent's age for users' convenience. In `BIOCOUPLY`, spell systems for each individual always start with the respondent's birth. The `SPELLTYP` of the first spell per definition is 'single'.

As `SPELLTYP` does contain missing values, so do `BEGIN(Y)` and `END(Y)`, indicating that the exact year of change in the couple status is not known. Missing dates indicate that the year was either not reported (-1) (or especially for 'married, separated'-spells the event of divorce is not reported) or that the reported year of change is implausible (-3), i.e. contradictory to other information. In order to differentiate the reasons for missing information the user can utilize the variables `REMARK` and `CENSOR`.

Consistent with `BIOCOUPLM` the data file `BIOCOUPLY` also provides the indicator variables `PDEATH` and `DIVORCE`. `PDEATH` indicates whether a respective spell ends with the death of a person's partner, but is not restricted to married persons, thus does not only refer to widowhood. The states of widowhood can easily be derived from `BIOCOUPLY` by using the spells of marriage in `SPELLTYP` and checking for the death of the respective spouse in `PDEATH` (also `BIOMARSM/Y` contains the marital status and therefore information about widowhood). `PDEATH` of single spells are assigned a '(-2) does not apply'. `DIVORCE` works in a similar fashion, indicating whether the last marriage spell, that is the separated spell, ended in divorce. Hence, if it did not end in divorce it is coded as a still ongoing marriage. In this case `END` is updated by the year of the last interview.

Variable `REMARK` provides information on whether we had to edit or supplement original information given by respondents in order to construct consistent couple biographies. Spells in `BIOCOUPLY` are marked as 'edited' (in contrast to 'original') if the editing process involved substitution of reported information because of inconsistency with responses in previous or following interviews or information reported by a partner. Furthermore we have added spells, e.g. between two marriages because two marriages with separate persons at the same time are not legal (see later section for more details on editing); these spells are marked as 'added spells'. Furthermore, 'first spells' and 'gap spells' are marked separately, whereas 'added' or 'edited' first spells are marked as 'first spells'. For `BIOCOUPLY` we also generated a variable `SOURCE`, which contains information about whether the respective spell was generated on basis of answers to the biography questionnaire or / and of responses to the personal questionnaire.

The variable `CENSOR` indicates whether a spell is left-censored, right-censored or censored on both sides (see Table 2 above within the explanations on `BIOCOUPLM`). The coding furthermore provides information on the reasons of censoring. In principle, spells might be censored if they precede or follow a gap spell or if `BEGIN` or `END` is missing. The last spell for each person is marked as right-censored if a person is still in the SOEP and the current relationship status is open ('last spell').

6 BIOMARSM: A monthly marital history

Spells in data file BIOMARSM contain prospectively collected information on marital biographies starting with the information reported in the first personal interview. The data file comprises eleven variables: the case and individual identifiers HHNR and PERSNR as well as nine spell specific variables. Variable SPELLTYP documents marital status with the possible categories ‘unmarried’, ‘married’, ‘divorced’, ‘widowed’ and ‘divorced or widowed’. Once married, a later spell ‘not married’ is not assigned anymore. Note that we renamed the known SOEP code ‘(1) single’ to ‘not married’. This is necessary to indicate that it is possible the respondent might have a partner anyway. If you are interested in this information, we recommend using BIOCPLM instead of BIOMARSM. Take also into account that we do not differentiate between marriages and registered same-sex partnerships for the SPELLTYP categories (3) to (6). SPELLTYP has one additional category ‘divorced or widowed’ which indicates that a marriage definitely ended though we do not know whether via divorce or death of the spouse. This may be due to missing information from the biographical questionnaires or due to a respondent’s frequent shifts between both categories without ever reporting the death of the partner or divorce as an event. A sixth state is ‘gap’ indicating a lack of reliable data for this period.

Variables of BIOMARSM

HHNR	Identifier of original sample household
PERSNR	Personal identifier
SPELLNR	Consecutive spell number (chronological order)
SPELLTYP	Marital status (1) Unmarried (2) Married (3) Divorced / reg. same-sex partnership annulled (4) Widowed / reg. same-sex partnership deceased (5) Divorced or widowed / reg. same-sex partnership annulled or deceased (6) Married, separated / reg. same-sex partnership, separated (7) Living in reg. same-sex partnership (9) Gap
BEGIN	Month when spell begins [1=Jan 1983 to 384=Dec 2014]
END	Month when spell ends [1=Jan 1983 to 384=Dec 2014]
BEGINY	Year spell begins [1983 to 2014; -3=implausible; -1=missing]
ENDY	Year spell begins [1983 to 2014; -3=implausible; -1=missing]
CENSOR	Censoring information [0 to 14] (see explanation above, Table 2)
EVENTS	Month information: exact month of spell begin or end known?
REMARK	Further spell information (1) Original spell (2) Edited spell (3) Added spell (4) Gap Spell (5) First Spell

Variable SPELLNR is a chronological index number for each individual's spells during the observation period. Variables BEGIN and END indicate the month in which a marital spell starts and ends. Monthly histories start with a value of 1 in January 1983 and ranges until the current margin in month 384, i.e., December 2014. In principle, the month in which a spell starts is the very same month the previous spell ends in. Please take into account that the months of BEGIN and END are not imputed when the exact month of the change in the familiar situation is not reported. Instead, in those cases the time of the interview is left as the BEGIN or END date when a change is observed in the data. To distinguish between these two cases we introduced a new variable called EVENTS, containing the information whether the exact month of the BEGIN or END of a spell is reported or not.

It is important to note that a 'first spell' in BIOMARSM is not the very first spell of a person, but the first observed marital status since the person is taking part in the SOEP. Accordingly, the first spell in BIOMARSM is almost always left-censored. Variable CENSOR informs about whether a spell is left- or right-censored and if so, why. Most spells in BIOMARSM are in fact censored. In order to provide the user with detailed information on the nature of censorship we distinguished 'left', 'right', and combined 'left- and right-censored spells' with respect to the reason for censoring: 'first spell' or 'last spell', 'spell ends with death', spell 'precedes' or 'succeeds a gap' (see Table 2 above within the explanations on BIOCOUNPLM). Of course, 'death' and 'last spell' are not mutually exclusive, thus we overwrite the latter with the former reason for being right-censored if the last interview is in the year of death or precedes it.

Variable REMARK provides information on whether we had to edit or supplement original information provided by respondents in order to construct consistent couple biographies. Spells in BIOMARSM are marked as 'edited' if the editing process involved substitution of reported information because of inconsistency with responses in previous or following interviews or information reported by a partner. Furthermore we have added spells, e.g. between two marriages because two marriages with separate persons at the same time are not legal (see later section for more details on editing); these spells are marked as 'added spells'. Finally 'first spells' and 'gap spells' are marked separately, whereas also 'added' or 'edited' first spells are marked as 'first spells'.

7 BIOMARSY: A annual marital biography

Data file BIOMARSY supplements BIOMARSM with retrospectively collected information on the marital history since a respondent's year of birth. Whereas the marital history in BIOMARSM is measured in months, BIOMARSY depicts the marital biography on an annual basis. In contrast to BIOCOUNPLY the BIOMARSY data set contains also the respondents to the biography questionnaire before wave 28. Please note, that until wave 28 the biography

questionnaire just asked for three previous marriages and therefore the number of reported marriages in BIOMARSY is limited.

The BIOMARSY file comprises eleven variables. The individual and household identifiers HHNR and PERSNR as well as SPELLTYP are basically the same in all data sets. Once married, a later spell ‘not married’ is not assigned anymore. Note again that we renamed the known SOEP code ‘(1) single’ to ‘not married’ and take into account that we do not differentiate between marriages and reg. same-sex partnerships for the SPELLTYP categories (3) to (6). This is to indicate that it is possible the respondent might have a partner anyway. Like BIOMARSM, it is important to notice that SPELLTYP has one additional category ‘divorced or widowed’ which indicates that a marriage definitely ended, though we do not know whether via divorce or death of the spouse. This may be due to missing information from the biographical questionnaires or due to a respondent’s frequent shifts between both categories without ever reporting the death of the partner or divorce as an event.

Variables of BIOMARSY

HHNR	Identifier of original sample household
PERSNR	Personal identifier
SPELLNR	Consecutive spell number (chronological order)
SPELLTYP	Marital status (1) Unmarried (2) Married (3) Divorced / reg. same-sex partnership annulled (4) Widowed / reg. same-sex partnership deceased (5) Divorced or widowed / reg. same-sex partnership annulled or deceased (6) Married, separated / reg. same-sex partnership, separated (7) Living in reg. same-sex partnership (9) Gap
BEGIN	Age of respondent when spell begins [-3=implausible; -1=missing]
END	Age of respondent when spell ends [-3=implausible; -1=missing]
BEGINY	Year when spell begins
ENDY	Year when spell ends
CENSOR	Censoring information [0 to 14] (see explanation above, Table 2)
SOURCE	Source of information (1) derived only from biography questionnaire (2) derived from biography and personal questionnaire (3) derived only from personal questionnaire
REMARK	Further spell information (1) Original spell (2) Edited spell (3) Added spell (4) Gap Spell (5) First Spell

Regarding the fact that duration of spells is measured in years it is important to notice that an individual may encounter several events in the same year. In this case the variable SPELLNR allows the user to order the spells with respect to a respondent’s life course. The variables BEGINY and ENDY provide the years in which a spell begins and ends, while the variables BEGIN and END indicate for users’ convenience the respective age of the respondent. The

spell system for each individual in BIOMARSY always starts with the birth of the respondents. We thus created a first spell for each individual ever interviewed in the SOEP starting in the year of birth and continuing at least until the year in which a person turns age 16. The SPELLTYP of the first spell per definition is ‘unmarried’. Even if a respondent reported an earlier marriage in the biography questionnaire we restricted its beginning to age 16, however, we marked the beginning of this spell as ‘implausible’.

There are some missing values (-1 and -3) in BEGINY as well as in ENDY (resp. BEGIN and END) indicating that we do not know the exact year of a change in the marital status. This can have two reasons. First, it may simply indicate that the respondent did not report the year in which a marriage began or ended. In order to differentiate the reasons for missing information the user may utilize the variables REMARK and CENSOR. Second, within single case corrections some dates are set ‘implausible (-3)’ if overlapping of marriages appear unsolvable or contradictions with the reported year and i.e. the year of death exist.

REMARK indicates whether a spell was ‘edited’ or ‘added’ in the same way as in BIOMARSM (see above). Variable CENSOR indicates if a spell is left or right censored or censored on both tails. The coding of CENSOR (see Table 2 above) provides also information about the reasons for censoring. In addition to what was said before about gap spells in BIOMARSM, gaps or missing values in BEGIN or END may appear in BIOMARSY if a respondent reported a terminated first marriage and the beginning of a second marriage, but did not report the reason for and/or the year of the end of the first marriage. The last spell of each person is marked as right censored, irrespective that the person died, quit the SOEP or is still married currently.