

Thorsten Euler/Maximilian Trommer

# For analysis of employment episodes in the Graduates Panel 2005

Problems and proposed solutions on the basis of  
employment episodes with multiple recording and survey  
overlapping duration

**Handout**

February 2019

The project on which this handbook is based was funded by the Federal Ministry of Education and Research under the number M521400. The German Centre for Higher Education Research and Science Studies GmbH is responsible for the content.

This work is licensed under the Creative Commons Attribution - Non Commercial - Distribution on the same terms 3.0 Germany License (CC-BY-NC-SA)

<https://creativecommons.org/licenses/by-nc-sa/3.0/de/>



Thorsten Euler

Telephone +49 (0)511 45 06 70-152 | Fax +49 (0)511 45 06 70-960

E-mail: [euler@dzhw.eu](mailto:euler@dzhw.eu)

Maximilian Trommer

Telephone +49 (0)511 45 06 70-148 | Fax +49 (0)511 45 06 70-960

E-mail: [trommer@dzhw.eu](mailto:trommer@dzhw.eu)

German Centre for Higher Education and Science Research GmbH

Lange Laube 12 | 30159 Hannover | [www.dzhw.eu](http://www.dzhw.eu)

February 2019

# 1 Introduction to the problem

The Graduate Panels of the DZHW have for a long time collected both employment episodes about periods between graduation and the respective survey dates as well as cross-sectional information about the current situation of the respondents at the time of the survey (e.g. current family circumstances) or information about the current or last employment. In the three waves of the 2005 graduate year cohort, gainful employment was recorded not only by means of a calendar (a mere indication of an occupation) but also by means of a tableau (variables `aocc22*`; `bocc22*` and `cocc22*`), in which additional information (e.g. the place of work) was asked for in addition to the start and end times for each employment episode.

In many cases, these employment episodes may have started during the observed period of the survey and not been completed. This applies to all current employment activities at the time of the first and second survey that still exist beyond the survey (see the example of the second survey (survey time/receipt of questionnaire in January 2011) in Figure 1; activity no. 2 of the second wave). The following section describes how the resulting problems and solutions were dealt with in the third survey of the 2005 graduate year. Unless otherwise stated, these steps were also applied in the second survey - deviations from these are listed in Chapter 4.

In the course of the third survey, the participants were again asked (after the two previous surveys) to report employment episodes and to accurately detail the beginning of each of them- even if this was in the period before the second survey time<sup>1</sup> (see example in Figure 1; Activities No. 1 and 2 of the third wave). As a result, double data are sometimes available for employment: an employment episode from the second survey that is marked as "still ongoing" at the time of the second wave and an employment episode from the third survey that begins before the second time of survey. Accordingly, the gainful employment between the beginning of the episode and the second time of the survey may be shown twice.

---

<sup>1</sup> Depending on the time of the response by the respondent(s) between winter 2010/11 and autumn 2011. Time of receipt of the questionnaire of the second survey according to the variable `bsys01*`.

## 2 Possible solutions

There were three possible solutions to this problem:

- a) Censoring: Termination of the current employment episode from the second survey (no "still running") and postponement of the start of the corresponding employment episode from the third survey to the time of the second survey.*

Advantage:

Both episodes would follow directly one after the other, duplication would be avoided.

Disadvantage:

The episodes seem to be separate, but refer to a single gainful activity. Accordingly, it is no longer possible to distinguish between a continuous gainful activity beyond the time of the survey and two independent gainful activities when changing at the time of the survey. In addition, an analysis of employment durations would underestimate the duration of employment, as there would be two short episodes instead of one longer one.

- b) Deletion: Deletion of the employment episode from the second survey and retention of the employment episode from the third survey with episode start before the second survey<sup>2</sup>.*

Advantage:

No duplication, the continuous episode would be entirely shown, an underestimation of the duration of employment would not be to be feared. Continuous employment episodes would differ from successive similar episodes.

Disadvantage:

The identification of the corresponding episodes is not always clearly possible due to deviating data (see chapter 3), so that there is a risk of permanently deleting one episode of the second survey by accident, although it is not a corresponding episode. In the case of corresponding episodes with different information for episode-related information, if an episode were deleted, there would also be the problem that this deviation (e.g. on the basis of promotion, task change, time limit removal, etc.) would no longer be visible.

---

<sup>2</sup> Alternatively, the reverse procedure (extension with extinction) with a temporal extension of the episode of the second wave to the known end of the episode of the third wave and subsequent extinction of the third wave episode would have been possible. Advantages and disadvantages for this variant would be the same as in solution b).

*c) Labelling: Retention of overlapping employment episodes and addition of a flag variable to identify the interrelated episodes.*

Advantage:

The data set reflects the above-mentioned problem of overlapping and possibly corresponding episodes of employment from the second and third surveys. The flag variable identifies the corresponding episodes and leaves it to the data user to decide how to deal with the problem, e.g. how to proceed in the case of solutions a) or b).

Disadvantage:

The data set is not ultimately cleansed, there are more episodes contained than were actually available. The flag variable is assigned for diverging information with a certain risk of error, which may not be taken into account by data users. If the employment episodes are not adjusted before analyses, there is a danger of a distorted representation (too many/short/running employment episodes).

## 3 Using a flag variable

### 3.1 Contents of the comparison figure

In the third survey of the Graduate Panel 2005, a decision was made in favour of the third solution for dealing with possibly related or corresponding episodes from the second and third surveys. The associated flag variable, which serves for the identification between employment episodes of the third survey and corresponding episodes of the second survey, is referred to below as the "comparison figure". Here, the comparison figure functions in relation to the previous survey - in the case of the third survey, the comparison with the second survey. The comparison figure can have a value of 0 (no corresponding episode in the previous wave) or 1-9 (corresponding episode in the previous wave). The system is based on the chronological order of the respective employment episodes in the pre-wave, as it already exists in the data set (bocc22\*). The value of the comparison figure indicates the position of the corresponding episode in the pre-wave. If, for example, an episode of the third wave has the comparison figure 4, it is assumed that this episode corresponds to the fourth employment episode of the second wave. Episodes of the third wave, which only began after the survey time of the second wave, generally received the comparison figure 0, since an overlap with pre-wave episodes is excluded in these cases.

### 3.2 Divergent episode information

An additional problem in identifying the "correspondence" of two episodes for a related gainful activity arises from the fact that the respondents' data do not match the episodically related information (employment, periods of employment, occupational status, etc.) for both episodes (although the employment is the same). There may be several reasons for this: on the one hand, the response given by respondents in the context of information retrieval varies to a certain extent, either due to difficulties in remembering or the divergent selection of the "applicable"<sup>3</sup> response category. On the other hand, within an employment, actual qualitative differences may emerge over time, resulting in a different answer to episode-related questions depending on the time of the answer<sup>4</sup>. Experience has shown that respondents do not always assess such changes within the framework of an existing employment as separate employment and (also for reasons of saving time when answering the questionnaire) do not reproduce them individually as individual employments (while retaining the remaining data). Accordingly, in the case of overlapping or corresponding episodes of the same employment, the respondents' episode-related information is often not completely identical.

### 3.3 Creation of comparison figure

A two-stage procedure was followed for the compilation of the comparison figures of the episodes of the third wave with episodes beginning before the second survey time. Firstly, an automated, syntax-based comparison of the employment episodes of the third survey with the employment episodes of the second survey was carried out. For each possible combination between an employment episode of the third wave and an employment episode of the previous wave, counter variables were generated which reflected the accordance of the episodes with the valid data on the temporal dimension (in particular the time of the beginning of the episode) and the employment-related information (on the basis of the employment, the working time model, the occupational position and the postcode of the place of work)<sup>5</sup>. On the basis of minimum thresholds for these matches, the most suitable combinations were determined and corresponding comparison figures were automatically assigned for the corresponding episodes. If the minimum threshold values were not reached, the comparison figure 0 was assigned conservatively, as correspondence of episodes was not sufficiently displayed and the possibility of an incorrect assignment was to be avoided.

---

<sup>3</sup> For example, a respondent might have classified himself/herself as an "executive employee" when questioning his/her occupational status, but instead have responded "qualified employee" when questioned again later, as there is a different subjective assessment of the qualification level of his/her own tasks.

<sup>4</sup> For example, an additional assumption of tasks in the course of employment may lead to a higher evaluation of the occupational position, the type of the employment contract may be, over time, shifted towards a permanent contract or the contractual working time may vary in the meantime.

<sup>5</sup> The contractual working hours and the federal state of the place of work were not compared in the employment-related information, among other things because the number of hours can change, for example, due to collective agreement conditions, and the federal state would remain constant even if the place of work was changed in the same federal state - however, the postal code can better display a change.

Figure 1: Example of employment episodes

Episode number	Beginning of episode		End of episode		Episode running?	Type of employment contract	Work time			Occupational status	Place of work		Comparison figure
												Postcode	
<i>1. Wave of employment episodes</i>													
1	Sep	2005	filters	filters	still running	Fixed-term	Part-time	no information		Scientific qualified employee(s) without management function	813	Bavaria	/
<i>2. Wave of employment episodes</i>													
1	Sep	2005	Sep	2008	not applicable	Fixed-term	Part-time	19 h		Scientific qualified employee without management function	813	Bavaria	1
2	Oct	2008	filters	filters	still running	Permanent	Full-time	39 h		Scientific qualified employee(s) with medium management function	353	Hesse	0
<i>3. Wave of employment episodes</i>													
1	Dec	2006	filters	filters	still running	other	Not agreed	no information		Qualified employee	no information	no information	0
2	Oct	2008	Jun	2011	not applicable	Permanent	Full-time	39 h		Scientific qualified employee without management function	no information	Hesse	2
3	Jun	2011	Jan	2014	not applicable	Permanent	Full-time	39 h		Scientific qualified employee without management function	315	Lower Saxony	0
4	Feb	2014	filters	filters	still running	no information	Full-time	39 h		Scientific qualified employee without management function	351	Lower Saxony	0

If several second wave episodes had the same fit (beyond the threshold), several third wave episodes matched the same pre-wave episode – which was especially the case if two similar employment episodes occurred simultaneously (e.g. two part-time positions with the same running time, work location and task level) - or if a third wave episode extended into the period of the second wave, but no corresponding episode could be automatically determined, the project team reviewed the individual cases. The latter occurred mainly in the case of third-wave episodes, which differed from the second-wave episodes in terms of employment-related information, but coincided with them in terms of time. On the basis of theoretical considerations on the above-mentioned problems of non-identical information from respondents or on labour market processes and changes within a continuous employment with the same employer, combination matrices of plausible deviations in the employment and the occupational position were generated (Figures 2 and 3).

Figure 2: Wave matrix of occupational status

3. Wave >		2. Wave:														
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Management employee	1	█														
Scientifically qualified employee, middle management	2		█													
Scientifically qualified employee, no management	3			█												
Qualified employee	4				█											
Executive employee	5					█										
Traineeship in a school, mandatory internship	6						█									
Self-employed, independent professionals	7							█								
Self-employed, independent entrepreneurs	8								█							
Self-employed (contract for work and services/fees)	9									█						
Public official in higher civil service	10										█					
Public official in higher intermediate service	11											█				
Public official in intermediate service	12												█			
Specialist (with apprenticeship)	13													█		
Unskilled/trained employee	14														█	
Assisting member of family	15															█



Figure 3: Wave matrix of employment contract

3. Wave >		Permanent contract	Fixed-term contract	Apprenticeship/Traineeship in a school	Contract for work and services/fees	Self-employed/freelance	Other
		1	2	3	4	5	6
2. Wave:							
Permanent contract	1						
Fixed-term contract	2						
Apprenticeship/Traineeship in a school	3						
Contract for work and services/fees	4						
Self-employed/freelance	5						
Other	6						

If plausible differences were found between second- and third-wave episodes (green highlighting) - e.g. through the transition from an apprenticeship to regular employment or from an executive to a qualified employee - the probability of correspondence increased. If there were implausible combinations - e.g. a change from a civil servant to an employee - the probability of correspondence decreased. In consideration of these individual case examinations, the still open or unclear comparison figures were assigned. If a better allocation of the episode correspondence in the individual case examination was not possible, the comparison figure 0 was also assigned conservatively. Figure 1 shows an example of the correspondence of two episodes. Figure 1 shows the employment episodes of all three waves in a fictitious case - episode no. 2 of the second wave continued ("still running") at the time of the survey. The starting time of episode no. 2 of the third wave is the same as that of the second wave, but both the information on the occupational position (with or without management function) and the postal code (missing information in the third wave) differ. Due to the plausible deviations in the occupational position (see Figure 2), the accordance with other episode-related information and the lack of other correspondence possibilities, the second episode of the third wave was assigned the comparison figure 2, for the other episodes of the third wave the comparison figure 0. The same applies to the comparison between episode 1 of the first wave and episode 1 of the second wave. Figure 4 illustrates the temporal distribution and overlapping of the episodes again visually.

## 4 Dealing with the episode connection problem in the second survey of the 2005 cohort

The function of the comparison figure as a flag variable was also used for the second survey of the 2005 Graduates Panel. The comparison figures of the episodes of the second wave (bocc22\*) identify there the corresponding employment episodes of the first wave (aocc22\*). Since the survey of the first wave allowed a maximum of six employment episodes for the period between completion of studies and the time of the survey, the comparison figures of the second wave showed either a value of 0 (no corresponding episode in the previous wave) or 1-6 (corresponding episode in the previous wave)<sup>6</sup>. In contrast to the third survey, the allocation of the comparison figures of the second survey was not automated, but exclusively allocated by means of individual case examination. The basis for the identification of the correspondence was the temporal dimension of the episodes (especially the time of the beginning of the episode) and the employment-related information (based on the employment, the working time model, the occupational position and the postal code of the place of work). Plausible deviations between the employment-related data were again<sup>7</sup> taken into account; implausible combinations were rejected as inadmissible.

## 5 Interpretation of the flag variable

The presented procedure of assigning comparison figures as flag variables to identify corresponding episodes of employment beyond the time of the survey is merely a guide for data users of the published data of the DZHW Graduate Panel. They serve for assignment of corresponding pre-wave episodes to follow-up episodes on the basis of the above criteria. However, these are only recommendations of the project team concerned with data collection and editing and entail a non-specifiable risk of incorrect allocation.

The possible adoption of these recommendations, an alternative identification of corresponding episodes, as well as the handling of overlapping employment episodes (see Chapter 2) are subject exclusively to the objectives, preferred analysis and choices of the respective data users.

---

<sup>6</sup> About 85 percent of those surveyed in the first survey reported at least one episode of employment after graduation, about 9 percent three or more episodes of employment. Accordingly, for most respondents the number of possible combinations between first and second wave episodes is rather low.

<sup>7</sup> The assignment of plausible combinations was not identical with the procedure in the third survey (see Figs. 2 and 3), but in accordance with it.

Figure 4: Visualization of the gainful activities in the case study (colouring according to Fig. 1) over time

(red = survey times wave 1, 2 and 3)

Wave		W1	W2		W3				Wave		W1	W2		W3			
Occupation		1	1	2	1	2	3	4	Occupation	1	1	2	1	2	3	4	
2005	Jan.								2011	Jan.							
	Feb.									Feb.			Still running				
	Mar.									Mar.							
	Apr.									Apr.							
	May									May							
	June									June							
	July									July							
	Aug.									Aug.							
	Sept.									Sept.							
	Oct.									Oct.							
	Nov.									Nov.							
	Dec.									Dec.							
2006	Jan.								2012	Jan.							
	Feb.									Feb.							
	Mar.									Mar.							
	Apr.	Still running								Apr.							
	May									May							
	June									June							
	July									July							
	Aug.									Aug.							
	Sept.									Sept.							
	Oct.									Oct.							
	Nov.									Nov.							
	Dec.									Dec.							
2007	Jan.								2013	Jan.							
	Feb.									Feb.							
	Mar.									Mar.							
	Apr.									Apr.							
	May									May							
	June									June							
	July									July							
	Aug.									Aug.							
	Sept.									Sept.							
	Oct.									Oct.							
	Nov.									Nov.							
	Dec.									Dec.							
2008	Jan.								2014	Jan.							
	Feb.									Feb.							
	Mar.									Mar.							
	Apr.									Apr.							
	May									May							
	June									June							
	July									July							
	Aug.									Aug.							
	Sept.									Sept.							
	Oct.									Oct.							
	Nov.									Nov.							
	Dec.									Dec.							
2009	Jan.								2015	Jan.							
	Feb.									Feb.							
	Mar.									Mar.							
	Apr.									Apr.							
	May									May							
	June									June							
	July									July							
	Aug.									Aug.							
	Sept.									Sept.							
	Oct.									Oct.							
	Nov.									Nov.							
	Dec.									Dec.							
2010	Jan.								2016	Jan.							
	Feb.									Feb.							
	Mar.									Mar.							
	Apr.									Apr.			Still running				Still running
	May									May							
	June									June							
	July									July							
	Aug.									Aug.							
	Sept.									Sept.							
	Oct.									Oct.							
	Nov.									Nov.							
	Dec.									Dec.							