

Anett Friedrich | Kevin Ord

# **BIBB Training Panel 2011 to 2017 Longitudinal data set**

Version 1.0





# DATA AND METHODOLOGICAL REPORT

No. 4/2020

Anett Friedrich | Kevin Ord

# BIBB Training Panel 2011 to 2017 Longitudinal data set

Version 1.0

## Acknowledgement

Special thanks for the support in the creation of the data set go to Christian Gerhards for the contribution in the harmonisation of the variable names and to Christine Hohn for the creation of the SPSS programme codes.

## Data availability

Access to the research data described in this handbook – as is the case with all firm-level data at BIBB-FDZ – is via the Remote Data Access and One-Site Visit.

## Note on the counting method for version numbers

Consecutive numbers after the comma (second level) document changes to the previous version without greater relevance to the content. On the other hand, changes relevant to the content lead to sequential numbering at the first level.

## Imprint

### Citation:

Friedrich, Anett; Ord, Kevin: BIBB Training Panel 2011 to 2017.  
Longitudinal data set. Data and methodological report 4/2020.  
Bonn 2020

1. Edition 2020

### Editor:

Federal Institute for Vocational Education and Training (BIBB)  
Robert-Schuman-Platz 3  
53175 Bonn  
www.bibb.de

### Publication management:

Strategic office "Publications and Scientific Information Services"  
E-Mail: publikationsmanagement@bibb.de  
www.bibb.de/veroeffentlichungen

### Production:

Verlag Barbara Budrich  
Stauffenbergstraße 7  
51379 Leverkusen  
Internet: www.budrich.de  
Email: info@budrich.de

### Contact FDZ:

Email: fdz@bibb.de  
Tel.: +49 228 107-2041  
Fax: +49 228 107-2020

### Licence:

The content of this work is subject to a Creative Commons Licence (licence type: Attribution – Non-Commercial – No Derivatives – 4.0 International).



For further information, please visit  
[www.bibb.de/oa](http://www.bibb.de/oa).

ISSN: 2190-300X

urn:nbn:de: 0035-0881-6

### Bibliographic information from the German National Library

The German National Library catalogues this publication in the German National Bibliography. Detailed information is available online at <http://dnb.dnb.de> abrufbar.

# Contents

<b>1</b>	<b>Introduction</b> . . . . .	<b>5</b>
1.1	Subject of the manual . . . . .	5
1.2	Structure of the manual . . . . .	5
1.3	Short description of the data set . . . . .	5
1.4	Central characteristics at a glance . . . . .	6
<b>2</b>	<b>Overview of the BIBB Training Panel</b> . . . . .	<b>7</b>
<b>3</b>	<b>Organisation of the data</b> . . . . .	<b>9</b>
3.1	Data sets . . . . .	9
3.2	Designation of the content variables . . . . .	12
3.3	Harmonisation of variables for longitudinal data sets . . . . .	13
<b>4</b>	<b>Organisation of microdata</b> . . . . .	<b>15</b>
4.1	Imputed variables . . . . .	15
4.2	Filters not recognisable in the questionnaire . . . . .	15
4.3	Missing values . . . . .	16
4.4	Number of employees in the establishment . . . . .	19
4.5	Special features of individual years . . . . .	19
4.6	English translation . . . . .	24
<b>5</b>	<b>Weighting and extrapolation</b> . . . . .	<b>25</b>
5.1	Calculation of weighting and extrapolation factors . . . . .	25
5.2	Weighting and extrapolation factors in the data set . . . . .	26
<b>6</b>	<b>Continuation in the following years</b> . . . . .	<b>30</b>
<b>7</b>	<b>Anonymisation and data access</b> . . . . .	<b>31</b>
7.1	Data protection provisions . . . . .	31
7.2	Formal data anonymisation . . . . .	31
7.3	Data access . . . . .	31
<b>8</b>	<b>References</b> . . . . .	<b>33</b>
	<b>Appendix</b> . . . . .	<b>34</b>



# 1 Introduction

## 1.1 Subject of the manual

This data and methodical report describes the longitudinal data set of the BIBB Establishment Panel on Qualification and Competence Development 2011 to 2017 (in short: BIBB Training Panel 2011 to 2017 long). The intention of the report is to help external researchers handling the data. Besides the documentation of the data structure and the harmonisation, the report contains Stata and SPSS commands that facilitate the work with the data and help to identify additional harmonisation options.

The data set is designed for longitudinal analyses specially. It makes sense to use the corresponding cross-sectional data set for analyses of individual waves of the BIBB Training Panel. The annual cross-sectional data sets are also available in the BIBB-FDZ.

Important core characteristics such as the response rate of the yearly surveys, the number of panel cases, as well as the research and education policy topics were bundled for all waves in the Excel file “Cross-sectional and panel characteristics of the BIBB Training Panel” by HOHN et al. 2020. The respective document is available on the BIBB-FDZ metadata portal page of the BIBB Training Panel.

## 1.2 Structure of the manual

Section 1.3 provides a short description of the data set and section 1.4 gives an overview of the central properties of the data set. Chapter 2 describes the BIBB Training Panel, explains the sampling concept and outlines the survey procedure. Following, Chapter 3 scrutinises the structure of the data sets, explains which subdata sets are available and what characterises them, and further outlines the harmonisation of variables over time. Afterwards, Chapter 4 explains the organisation of the microdata, in particular missing values, imputation, filters not recognisable in the questionnaire and special features of individual years and the possibility of translating the data set into English. Chapter 5 deals with weighting and extrapolation factors. Finally, Chapter 6 provides information on the updating of the long data set and Chapter 7 on the anonymisation and data access.

## 1.3 Short description of the data set

The longitudinal data set of the BIBB Establishment Panel on Qualification and Competence Development 2011 to 2017 comprises one main data set and ten supplementary subdata sets (for details see Section 3). The main data set contains the variables regularly surveyed between 2011 and 2017. A first subdata set contains sporadic surveyed variables, a second subdata set contains the weighting and extrapolation factors. The remaining eight subdata sets<sup>1</sup> contain the variables that were only queried in the corresponding survey year. The total data set includes 24,075 observations from 13,473 establishments and 1,172 variables.

---

<sup>1</sup> For the year 2016 two subdata sets exist because an additional CATI survey was conducted in this year.

## 1.4 Central characteristics at a glance

Survey title	<b>BIBB Establishment Panel on Qualification and Competence Development 2011 to 2017 long</b>
DOI	10.7803/371.1117.1.2.10
Brief description	Establishment panel with the aim of providing detailed information on the structures, developments and interrelationships of in-establishment training measures and on the demand for skilled labour in establishments
Survey year	2011 to 2017
Survey unit	Establishments
Thematic topic	Vocational education and training, continuing training
Data access options	On-site use/Remote data access
Number of variables	1,172
Population	All establishments in Germany with at least one employee subject to social insurance contributions
Weighting/Extrapolation	Extrapolation factors (based on the establishment register of the Federal Employment Agency (BA)), Longitudinal and cross-sectional weights
Representative region	East and West Germany
Number of cases	24,075 observations from 13,473 establishments
Survey procedure	CAPI (PAPI or CAWI; at the request of the establishment)
Selection procedure	Disproportionately stratified random selection
Survey design	Longitudinal design (panel study)
Remark	The data set is divided into eleven subdata sets. The main data set is stored in long format.
Links	BIBB-FDZ Metadata Portal: <a href="https://metadaten.bibb.de/metadaten-gruppe/16">https://metadaten.bibb.de/metadaten-gruppe/16</a> Project page: <a href="https://www.bibb.de/en/1482.php">https://www.bibb.de/en/1482.php</a>
Keywords	Qualification, vocational education and training (VET), staff movements, work organisation, staff structure, VET of refugees, technological change, job tasks, termination of VET contracts, organisation and implementation of continuing training



## 2 Overview of the BIBB Training Panel

The BIBB Training Panel, which has been conducted annually since 2011, is a representative panel survey of about 2,000 and from 2014 of about 3,500 establishments (cf. HOHN et al. 2020) in Germany. The focus of the study is on establishment structure and developments with regard to vocational education and training (VET), continuing training, staff structure and work organisation within an establishment. However, in all waves of the survey, other key topics, such as establishment digitisation or VET and employment of people with a refugee background<sup>2</sup>, supplement the base information.

In the sixth wave (2016) of the Training Panel parallel to the main study, the additional thematic focus was surveyed in Computer Assisted Telephone Interviews (CATI). 3,500 additional establishments, hence in total of about 7,000 establishments, took part in survey in 2016.

The population of the BIBB Training Panel comprises all establishments in Germany with at least one employee subject to social insurance contributions as of 31 December of the previous year of the survey. An establishment is defined as an economically and regionally separate unit with at least one employee subjected to social insurance contributions (cf. FISCHER et al. 2008: 7; GERHARDS et al. 2012). The sample, or more precisely the first gross sample, is drawn from the establishment register of the Federal Employment Agency (BA). The net sample is divided into re-surveyed (panel) establishments and refresher establishments, which are needed to compensate for the panel mortality. In 2017, only addresses of the establishments, which in the aforementioned separate CATI survey from the 2016 wave expressed their willingness to participate again, were used for the refresher sample.

The gross sample, i.e. the sample consisting of all the contacted establishments, is disproportionately stratified according to four criteria:

- ▶ the regional distribution of the establishments according to East and West Germany,
- ▶ the size of the establishments, divided into four classes (1-19 employees, 20-99 employees, 100-199 employees and 200 and more employees),
- ▶ the industry, divided into eight classes (agriculture/mining and energy; manufacturing; construction; trade and repair; business services; personal services; medical services; public services and education),
- ▶ whether the establishment provides VET.

The sampling of the BIBB Training Panel is disproportionately stratified, since otherwise the number of cases may be too small to conduct meaningful statistical analyses. Specifically, the proportion of establishments providing VET, establishments in eastern Germany, establishments from certain industries, such as manufacturing, and establishments with at least 200 employees subject to social insurance contributions in the sample is higher than the proportion of these groups in the population.

The surveys are carried out as Computer Assisted Personal Interviews (CAPI). Since 2016, establishments that do not wish to conduct a personal interview get the opportunity to participate via Computer Assisted Web Interview (CAWI) with identical questions. Until 2015, instead of CAWI, a Paper and Pencil Interview (PAPI) was offered as an alternative to CAPI.

---

2 For a list of all thematic focuses, see HOHN et al. 2020.

Before the field phase, the entire questionnaire is tested in a pre-test with about 30 establishments for comprehension difficulties, length, complexity and programming errors. The BIBB tests the resulting final version of the CAPI and CAWI questionnaire before its usage in the field. The field phase usually starts at the beginning of February and runs until the beginning of September.

Until 2014, “TNS Infratest Sozialforschung”<sup>3</sup> and from 2015 “infas Institut für angewandte Sozialwissenschaft” conducted the interviews of the BIBB Training Panel. The survey institute sends previously trained interviewers to the participating establishments to conduct the interviews on site. As interview partners, persons with a sufficient overview of the internal activities of the establishment are recruited. Typically, these are owners, shareholders or the managers of the respective human resource department. The choice of the interviewee within the establishment is left to the establishment.

The field phase starts with the written contact of the establishments from the gross sample. A letter informs about the start of the next survey wave and the current thematic focus and asks to participate. In addition, the letter announces the forthcoming contact of the establishment via an interviewer. The letter also contains a data protection declaration from the BIBB and the survey institute conducting the survey. Since 2016, the establishments in the refresher sample and establishments for which no contact person was known have been contacted in advance by telephone.

In the course of the contact by the interviewers, the latter will resend the letter to the establishments on request. If necessary, an overview of the questions is also handed over so that the respective interview partner can prepare sufficiently for the interview. If the establishment refuses to participate in CAPI, the interviewers offer to participate in the CAWI version of the survey (in previous waves PAPI).

On average, between 2011 and 2017, the BIBB Training Panel had a response rate<sup>4</sup> of around 36.8% and a willingness of establishments to participate in the panel of around 85% (HOHN et al. 2020).

---

<sup>3</sup> TNS Infratest Social Research has now changed its name to “Kantar TNS”.

<sup>4</sup> The BIBB Qualification Panel defines response as the number of interviews that can be evaluated divided by the number of addresses used.

## 3 Organisation of the data

### 3.1 Data sets

The BIBB Training Panel 2011 to 2017 long comprises a total of eleven subdata sets (see Table 1 below), two longitudinal data sets, one data set with the weighting and extrapolation factors and one cross-sectional data set per survey year. The year 2016 is an exception; due to the additional CATI survey there are two cross-sectional data sets for this survey year.

All variables that are included at least between 2015 and 2017 or are permanently queried from 2016 onwards are contained in the data set “Qualifizierungspanel\_11bis17\_long\_gwa”<sup>5</sup>. All other variables that do not meet these criteria but have been surveyed at least twice are contained in the data set “Qualifizierungspanel\_11bis17\_long2\_gwa”. All variables collected only once are included in an annual data set “Qualifizierungspanel\_11bis17\_quer\_20\*\_gwa” (\*=11 to 17).

**Table 1: Overview of partial subdata sets**

	Data set	Content	Variables (columns)	Observations (rows)	Establishments
1	Qualifizierungspanel_11bis17_long_gwa	<b>Main data set</b> , Longitudinal data set, All variables included at least between 2015 and 2017, or permanently queried from 2016 onwards	260	24,075	13,473
2	Qualifizierungspanel_11bis17_long2_gwa	Longitudinal data set, All other variables not meeting these criteria but collected at least twice	122	24,075	13,473
3	Qualifizierungspanel_11bis17_weight_gwa	Weighting and extrapolation factors	27	24,075	13,473
4	Qualifizierungspanel_11bis17_quer_2011_gwa	Cross-sectional data set 2011	137	2,004	2,004
5	Qualifizierungspanel_11bis17_quer_2012_gwa	Cross-sectional data set 2012	118	2,023	2,023
6	Qualifizierungspanel_11bis17_quer_2013_gwa	Cross-sectional data set 2013	102	2,063	2,063
7	Qualifizierungspanel_11bis17_quer_2014_gwa	Cross-sectional data set 2014	40	3,523	3,523
8	Qualifizierungspanel_11bis17_quer_2015_gwa	Cross-sectional data set 2015	67	3,589	3,589
9	Qualifizierungspanel_11bis17_quer_2016_gwa	Cross-sectional data set 2016_gwa	40	3,616	3,616
10	Qualifizierungspanel_11bis17_quer_2016_CATI_gwa	Cross-sectional data set 2016 CATI supplementary survey	45	3,521	3,521
11	Qualifizierungspanel_11bis17_quer_2017_gwa	Cross-sectional data set 2017_gwa	213	3,727	3,727

<sup>5</sup> There are four exceptions where the variables are included in the long2 data set, although they were queried exactly between 2015 and 2017. The reason is that the variables will no longer be collected from 2018 onwards.

The main data set “Qualifizierungspanel\_11bis17\_long\_gwa” forms the core element of the BIBB Training Panel 2011 to 2017 long. The data set is stored in the long format, which means that establishments are included in the data set as often as they participated in the survey. By contrast, each variable (e.g. well (wave indicator)) is only available once, i.e. the information is stored row by row<sup>6</sup>.

In total, the main data set contains 260 variables and 24,075 observations from 13,473 different establishments. Table 2 shows the number of waves in which the establishments participated. On average, the establishments participated in 2.5 surveys, with around a half participating only once and around one percent in all seven waves.

**Table 2: Frequency of establishment participation in the BIBB Training Panel between 2011 and 2017**

Total number of participations until 2017	Absolute number	Percent
1	7,471	55.45
2	3,390	25.16
3	1,496	11.10
4	670	4.97
5	170	1.26
6	126	0.94
7	150	1.11
<b>Total (number of establishments)</b>	<b>13,473</b>	<b>100.00</b>

Source: BIBB Training Panel 2011 to 2017 long

In total, 71 patterns for the participation of establishment exist; in Stata you can view them with the command `xtides, pattern(71)`. Table 3 shows common patterns and how many establishments participated in the corresponding period. For example, 2,097 establishments participated only between 2016 and 2017, but if all other possible patterns are included, 3,607 establishments are available for the evaluation of the two years.

**Table 3: Pattern of establishment participation in the BIBB Training Panel between 2011 and 2017**

Time	Number of establishments participating only during this period	Total number of participating establishments during this period
2016-2017	2,097	3,607
2015-2017	906	1,496
2015-2016	302	1,969
2014-2017	294	582
2011-2014	287	561
2011-2013	285	858
2011-2017	150	150

Source: BIBB Training Panel 2011 to 2017 long

The main data set contains, as already mentioned, the panel variables, which are regularly collected. A specially created questionnaire contains the respective panel variables and is avail-

<sup>6</sup> In the wide format, however, each establishment would exist only once, and each variable would exist once for each survey year with correspondingly different names, e.g. well2011, well2012...well2017 (i.e. the information would be stored in columns).

able at the BIBB-FDZ website. However, even the variables in the long data set were not surveyed consistently in all years. For example, questions on refugees or on technological change were not asked until 2015 or 2016. Further, the number of continuing training participants in 2011 and 2013 was not surveyed separately for employment groups. Variables and items not available for all years can be found in the Appendix Table A1.

The second long data set “Qualifizierungspanel\_11bis17\_long2\_gwa” comprises 122 variables and 24,075 observations from 13,473 different establishments, too. This data set contains all variables that were surveyed more than once but not on a permanent basis, such as questions on employees tasks or the perceptual performance level of apprentices. The exact variables contained in the data set and the year in which they were surveyed can be seen in the list of variables in the Appendix Table A2. The data set is also stored in long format.

The data set “Qualifizierungspanel\_11bis17\_weight\_gwa” contains 27 variables and the same number of observations and establishments as the two other long data sets. It is also stored in long format. In chapter 5 “Weighting and extrapolation” the data set is described in more detail.

The cross-sectional data sets contain all variables that were collected once in a given year or that were asked in a specific way. The number of variables tends to decrease over time (see Table 1) as the set of panel questions has become more stable and extensive over the years. The cross-sectional data sets also contain cases that actually do not belong to the cross section because they are not part of the population or because the establishment has become extinct. These establishments can be identified by the fact that they contain the value -11 ‘Filter (establishment extinct)’<sup>7</sup> for all content variables. For an overview, lists of the cross-sectional variables are located in the BIBB-FDZ metadata portal (available in German only). The introduced cross-sectional data sets were created specifically for merging variables to the longitudinal data sets. The initial cross-sectional data sets previously published separately in the BIBB-FDZ should be used to evaluate the cross section of individual years. These data sets have the advantage that they contain all the variables collected in a given year for establishments participating in that year.

In the long data sets, the variables *id2* and *year* define a unique data row. These two variables must be used to merge all data sets. The corresponding commands are:

```
Stata
use "Qualifizierungspanel_11bis17_long_gwa.dta"
sort id2 year
merge 1:1 id2 year using "Qualifizierungspanel_11bis17_long2_gwa.dta", keepusing(varlist)
cap drop _merge
merge 1:n id2 year using "Qualifizierungspanel_11bis17_quer_2011_gwa.dta", keepusing(varlist)
.
.
.
cap drop _merge
merge 1:n id2 year using "Qualifizierungspanel_11bis17_quer_2017_gwa.dta", keepusing(varlist)
```

7 In the data set ‘Qualifizierungspanel\_11bis17\_quer\_2012\_gwa’ the value ‘-1111. filter (establishment extinct)’ is for the variables *b43\*p\_gen*, as -11 could theoretically be a valid value.

```
cap drop _merge
merge 1:1 id2 year using "Qualifizierungspanel_11bis17_weight_
gwa.dta", keepusing(varlist)
cap drop _merge
```

### SPSS

```
GET FILE="*\Qualifizierungspanel_11bis17_long_gwa.sav".
SORT CASES BY id2 year.
MATCH FILES /FILE=*
  /FILE='*\Qualifizierungspanel_11bis17_long2_gwa.sav'
  /RENAME (pan_ges transverse well = d0 d1 d2)
  /BY id2 year
  /DROP= d0 d1 d2.
EXECUTE.
MATCH FILES /FILE=*
  /TABLE='*\Qualifizierungspanel_11bis17_quer_2011_gwa.sav'
  .
  .
  .
  /TABLE='*\Qualifizierungspanel_11bis17_quer_2017_gwa.sav'
  /FILE='*\Qualifizierungspanel_11bis17_weight_gwa.sav'
  /BY id2 year.
EXECUTE.
```

## 3.2 Designation of the content variables

Until 2015, the naming of the content variables follows the numbering of questions from the questionnaire. The variable names of the content variables begin with “a” for wave 1, “b” for wave 2 to “e” for wave 5. The two-digit question number of the questionnaire follows the wave indicator. An addition (e.g. “a01a” and “a01b”) complements the variable name for sub-questions. From 2016, a stable variable naming system replaced the previously used naming of variables. Now each variable consists of two to three parts:

1. A two-letter prefix indicating the subject area to which the variable can be assigned (cf. Table 4);
2. A consecutive number;
3. If necessary, a suffix, which is an abbreviation used in the previous waves (cf. HOHN et al. 2020).

**Table 4:** System of variable names

Variable prefixes	Subject area
ab	VET
ps	Staff structure
pb	Staff movement
wb	Continuing training
mo	Special modules
di	Technology and automation
be	General information about the establishment

The annual cross-sectional data sets before 2016 retain the original designation if the question was asked only once. In contrast, variables were renamed retroactively if the question could be assigned to the new logic. This is also the case when the whole question is not assignable to the new one, but a single item from a battery can be assigned to other items from other waves. For an easier traceability in the original questionnaires, the question number was added to the variable label. The following two examples from the cross-sectional data set 2011 illustrate the naming scheme and the harmonisation of variable names (cf. Table 5).

**Table 5: Example of the system of variable names in a cross section**

Variable	Variable label	Explanation
a02	Need for further qualification of employees	Question was only asked in 2011; it follows the scheme a= wave2011 (b=20112).
ab058_4	@A03d: VET causes high costs	Question ab058 was also asked in other waves; it follows the scheme "ab"= question on the subject of VET. Item _4 is included in the cross section because it was only asked once in the wave 2011. The naming with the prefix ab058 makes it easier to find related items over several waves.

### 3.3 Harmonisation of variables for longitudinal data sets

To ensure a uniform naming of the variables in the long data sets, the variable names of the waves before 2016 were adapted to the current naming system. For an easier traceability of variables before 2016 in the original questionnaires, an allocation table is available on the pages of the BIBB-FDZ metadata portal for each of the variables in the long and long2 data sets (in German only).

Furthermore, in all variables, cases were coded as 9997, 997 or 97 if the corresponding question was not asked in the respective wave and as 9996 or 96 (in one case 95) if the corresponding item was not asked. In addition, in all imputed variables, cases were coded in 998, 9998, 99998 or -98<sup>8</sup> if there was no imputation of these variables in the respective wave. These categories should facilitate the orientation in an initial exploration of the data. For analysing the data, it may be useful to define these values as System Missings; for a programming aid see chapter 4.3.

In the years 2012 to 2014, the cross sections included establishments that were already extinct but were by definition still part of the panel and hence were included in the weighting (cf. GERHARDS/FRIEDRICH 2016). The respective establishments are marked with '-11. filter (establishment extinct)' if the corresponding variable is filled for the respective year. Starting with the 2015 wave, the definition of panel establishments changed. As a result, extinct establishments are no longer panel establishments and are no longer included in the calculation of the weights (cf. GERHARDS/FRIEDRICH 2017).

In some cases, there were slight changes in the formulation of panel questions. Corresponding modifications are documented in the notes (for Stata, in German only) (commands for displaying the notes: `notes varname`) and can also be found in HOHN et al. 2020.

Table A3 in the Appendix documents the detailed harmonisation of individual variables of the two long data. The variables have been modified in such a way that they are uniform in the long data sets for all years despite annual differences in the survey (e.g. different number of

<sup>8</sup> A negative value was coded for the business volume because a value that would be high enough to be the highest value cannot be labelled for system-related reasons.

response categories or separate collection of data for industrial-technical and commercial-administrative apprentices).

The variables of the cross sections have not been harmonised. An exception is the 2011 variable *wb011\**, where the category '9999. group of employees does not exist' was coded.

In addition to the harmonisation listed above, the internal research data of the BIBB Training Panel was adjusted in comparison to the cross-sectional data sets published so far. These adjustments involve corrections or updates to the data that were made by the project team in the course of the project work. They are included in the data but not listed in detail.



## 4 Organisation of microdata

### 4.1 Imputed variables

The BIBB Training Panel only imputes continuous variables with a high item non-response rate (generally over 10%). All cases were imputed that either did not provide any information or responded “don’t know”. Missings or extreme values were not imputed. The aim of imputation is on the one hand to mitigate the bias caused by item non-response and on the other hand to increase the number of cases for analysis.

A linear prediction value is calculated using a common regression model, which takes into account, among other things, important structural variables. Panel information, i.e. values from previous years, are partly used for imputation. The imputed value is the linear prediction value of the regression model with the addition of an error term with 1,000 repetitions. For a detailed description of the imputation procedure, see GERHARDS/FRIEDRICH 2016, Chapter 4.

The imputed variables are each marked with the abbreviation “\_imp” and are included in the data set in addition to the original variables. It is up to each user to decide whether to use the imputed or the original variable. Since metric variables with a high item non-response are imputed instead of a fixed set of variables, it is not constant which variables have been imputed. Once a variable has been imputed for a year, it is included in the long or long2 data set.

### 4.2 Filters not recognisable in the questionnaire

For some variables, filters not visible in the questionnaire were coded in the respective year when the cross sections were prepared. There are two reasons for this:

Firstly, concerning the question about VET occupations and the number of apprentices in these occupations, it is useful to distinguish between establishments without apprentices and those that provide VET in less than five occupations. To make it easier to distinguish these cases from the other missing values, they were coded as ‘99999. No further VET occupation’.

Secondly, in the case of questions relating to specific groups of employees, it makes sense to distinguish the establishments which were not asked the question due to the absence of the relevant group of employees from those establishments which were not asked the questionnaire due to other filters. The filtering is based on the number of employees in the respective employee group (*ps20\**) and the number of employees of respective employee group who left the establishment (*pb017\**). The commands for tracing the filters are:

#### Stata

```
gen filter_gering=.
replace filter_gering=1 if ps020_enf>0 | pb017_enf>0 & pb017_enf<
9997
replace filter_gering=0 if ps020_einf==0 & pb017_einf==0

gen filter_qual=.
replace filter_qual=1 if ps020_qual>0 | pb017_qual>0 & pb017_
einf<9997
replace filter_qual=0 if ps020_qual==0 & pb017_qual==0
```

```

gen filter_high=.
replace filter_high=1 if ps020_high>0 | pb017_high>0 & pb017_f<
9997
replace filter_high=0 if ps020_high==0 & pb017_high==0

```

**SPSS**

```

compute filter_gering=$Sysmis.
if ps020_ins>0 | pb017_ins>0 & pb017_ins<9997 filter_gering=1.
if ps020_ins=0 & pb017_ins=0 filter_gering=0.

compute filter_qual=$Sysmis.
if ps020_qual>0 | pb017_qual>0 & pb017_einf<9997 filter_qual=1.
if ps020_qual=0 & pb017_qual=0 filter_qual=0.

compute filter_high=$Sysmis.
if ps020_high>0 | pb017_high>0 & pb017_f<9997 filter_high=1.
if ps020_high=0 & pb017_high=0 filter_high=0.
EXECUTE.

```

All cases that do not have a value of 1 in the corresponding variables were assigned the codes 9999, 999 or 99 ‘Employee group not exists’. An exception in the coding is the year 2012, in which the information on employees who have left the establishment is missing (cf. Chapter 4.5.2).

### 4.3 Missing values

Missing and invalid values were recoded in accordance with the standard declaration of missing values of the BIBB-FDZ. Special features of the coding from previous waves were taken into account for all variables of the BIBB Training Panel. HORN et al. 2020 (Table sheet “Missing values 13”) contains the values and value labels for the missing values of all variables of the BIBB Training Panel.

The BIBB-FDZ generally defines no System Missings in its data sets; it is the users’ decision to assign them. Exceptions are the weighting and extrapolation factors, which contain System Missings.

After the initial exploration of the data, it may be useful for analyses to define System Missings for the categories: missing values, questions not asked in the respective year, imputations not carried out and filters not recognisable in the questionnaire. The corresponding commands are:

**Stata**

```

*****long
***Question not asked in the corresponding year
mvdecode ab002 ab036 ab057 ab065 ab069 ps036 pb001_ang wb014
wb017_* di001_* be001_* be010_* be043 be026*k ,mv(97=.a)
mvdecode ab034 ab037 ab067 ab068 ps037 wb003*, mv(997=.a)
mvdecode ab023_* ps001aus ps001soz ps001ger_v ps001aus_v
ps001soz_v pb005* pb011* pb017* wb002* wb010*, mv(9997=.a)
mvdecode ps001zwi ps001bea ps001svb_v ps001nsv_v ps001zwi_v

```

```

ps001bea_v ps030_sch ps030_sch_imp, mv(99997=.a)
***No imputation in the corresponding year
mvdecode ab034_imp ab037_imp wb003*_imp , mv(998=.b)
mvdecode ab017_imp ab018*_imp ab023*_imp ab033_imp ab039*_
imp ab043*_imp pb003_imp pb016_imp ps015_2_imp pb003_f_imp
pb005*_imp pb010_imp pb011*_imp pb017*_imp wb002_imp wb010*_
imp,mv(9998=.b)
mvdecode ab004_b*_imp ps001ger_imp ps001nsv_imp ps001soz_imp
ps001zwi_imp ps001svb_v_imp ps001ger_v_imp ps001nsv_v_imp
ps017*_imp ps020*_imp ps030_uni_imp wb009_imp,mv(99998=.b)
mvdecode be026_* ,mv(999998=.b)
mvdecode be023_imp ,mv(-98=.b)
***Employee group not exists
mvdecode be026_*k, mv(99=.c)
mvdecode pb003_f pb005* wb010*, mv(9999=.c)
mvdecode be026_*, mv(999999=.c)
***No further VET occupation
mvdecode ab004* kldb2010_5er ab004_b* ,mv(99999=.d)
mvdecode ab004*pr_dl ,mv(99=.d)
***Item not queried
mvdecode ab004*pr_dl ab058_5 ab058_6 ab058_7 wb017_4 be001_6
be001_7 be001_8 be001_9 be010_7 be010_8 be035_3 be035_4 be035_5
be035_6,mv(96=.e)
mvdecode ab043_*, mv(9996=.e)
***Continuing training participation cases
mvdecode wb010*,mv(9995=.f)
***Missings
mvdecode *, mv(-12/-1)

*****long2
***Question not asked in the corresponding year
mvdecode ab009 ab027* ab035 ab056 ab064 ab083 ab085* ab089
wb015_1 wb016_1 wb017_* wb018 di018* mo001* be022 be042_*,
mv(97=.a)
mvdecode ab055* ab083*ab086* ab087* wb015_2* wb016_*,
mv(997=.a)
mvdecode ab003* ab082* ps039*, mv(9997=.a)
***No imputation in the corresponding year
mvdecode ab083*_imp ,mv(998=.b)
mvdecode ab003*imp ab082*imp, mv(9998=.b)
***Employee group not exists
mvdecode wb017* mo*,mv(99=.c)
mvdecode ps039*, mv(9999=.c)
***Nor further VET occupation
mvdecode ab086* ab087*, mv(999=.d)
***Item not queried
mvdecode ab027_2 ab027_5 ab027_6 , mv(96=.e)
mvdecode ab027_7 ab027_8, mv(95=.e)
***Missings
mvdecode *,mv(-12/-1)

```

**SPSS**

```

*****long
***Question not asked in the corresponding year.
missing values ab002 ab036 ab057 ab065 ab069 ps036 pb001_ang
wb014 wb001 wb017_1 to wb017_5 di001_2 to di001_8 be001_1 to
be001_9 be010_1 to be010_8 be043 be026_einfk to be026_hochk (97)
ab034 ab037 ab067 ab068 ps037 wb003kv to wb003so_imp (997)
ab023_1 to ab023_4_imp ps001aus ps001soz ps001ger_v ps001aus_v
ps001soz_v pb005* pb011_einf to pb011_hoch_imp pb017_einf to
pb017_hoch_imp wb002 wb002_imp wb010einf to wb010hoch_imp
(9997).
***No imputation in the corresponding year.
missing values ab034_imp ab037_imp wb003gt_imp to wb003so_imp
(998)
be023_imp (-98)
ab017_imp ab018_1_imp to ab018_4_imp ab023_1_imp to ab023_4_imp
ab033_imp ab039_gt_imp ab039_kv_imp ab043_gt1_imp to ab043_kv3_
imp pb003_imp pb016_imp ps015_2_imp pb003_f_imp
pb005_einf_imp to pb005_hoch_imp pb010_imp pb011_einf_imp to
pb011_hoch_imp pb017_einf_imp to pb017_hoch_imp wb002_imp
wb010einf_imp to wb010hoch_imp (9998).
***Employee group not exists.
missing values be026_einfk to be026_hochk (99)
pb003_f pb005_into pb005_high_imp wb010into wb010high_imp ab004_
b2 to ab004_b5_imp (9999).
***No further VET occupation.
MISSING VALUES ab004_a1_kldb2010_5er to ab004_b5_imp (99999).
***Item not queried.
MISSING VALUES ab004*pr_dl ab058_5 ab058_6 ab058_7 wb017_4
be001_6 be001_7 be001_8 be001_9 be010_7 be010_8 be035_3 be035_4
be035_5 be035_6 (96)
ab043_gt1 to ab043_kv3_imp (9996).
***Continuing training participation cases
MISSING VALUES wb010* (9995).
***Missings.
MISSING VALUES all (-12 thru-1).

*****long2.
***Question not asked in the corresponding year.
MISSING VALUES ab009 ab027_1 to ab027_8 ab035 ab056 ab064 ab083
ab085gt ab085kv ab089 wb015_1 wb016_1 wb017_einf1 to wb017_hoch5
wb018
di018_1 to di018_6 mo001einf_1 to mo001hoch_8 be022 be042_1 to
be042_3 (97)
ab055 ab055_imp ab083 to ab083kv_imp ab086gt1 to ab086kv4
ab087gt1 to ab087kv4 wb015_2 wb015_2_imp wb016_1 to wb016_2_imp
(997)
ab003f to ab003kv_imp ab082gt1 to ab082kv4_imp ps039 ps039_imp
(9997).

```

```

***No imputation in the corresponding year.
MISSING VALUES ab083gt_imp ab083kv_imp (998)
  ab003f_imp ab003gt_imp ab003kv_imp ab082gt1_imp to ab082kv4_imp
(9998) .
***Employee group not exists.
MISSING VALUES wb017_einfl to wb017_hoch5 mo001einf_1 to
mo001hoch_8 (99)
  ps039 ps039_imp (9999) .
***No further VET occupation.
MISSING VALUES ab086gt1 to ab086kv4 ab087gt1 to ab087kv4 (999) .
***Item not queried.
MISSING VALUES ab027_2 ab027_5 ab027_6 (96)
  ab027_7 ab027_8 (95) .
***Missings.
MISSING VALUES all (-12 thru -1) .

```

#### 4.4 Number of employees in the establishment

The BIBB Training Panel collects various data on employees, namely:

- ▶ employees subject to social security contributions,
- ▶ marginally employed persons,
- ▶ active owners and assisting family members,
- ▶ apprentices according to BBiG/HwO,
- ▶ apprentices in the health sector,
- ▶ candidates for civil servants.

In addition, there is variable *ps001zwi*, which contains the sum of employees subject to social security contributions and marginally employed persons and active owners/assisting family members. This variable is suitable as a basis for calculating shares for variables such as *ps017\** (number of women, part-time and fixed-term employees), *ps020\** (number by group of employees), *ps030\** (number by level of education), *ps037* (number of refugees), *wb002* and *wb003* (number of courses for advanced VET), as well as *wb009* and *wb010\** (number of continuing training courses). In individual cases, however, values above 1 may occur.

#### 4.5 Special features of individual years

In individual waves, variations in the meaning of certain variables or differences in their collection mode exist. The following chapters provide information on the corresponding particularities and the resulting consequences.

##### 4.5.1 2011

Concerning the scope of continuing training in 2011, the number of participants (persons) or the cases of participation (*wb011\**) were queried. In contrast, in the other years, only the number of continuing training participants (*wb010\**) was queried. By means of variable *a44*, it is possible to distinguish whether participants or participation cases were reported in 2011.

The information on participants for 2011 has been included in variable *wb010\** (cf. Appendix Table A3), whereas the information on participation cases has not been included in *wb010\**.

Transferring all values from *wb011\** to *wb010\** would overestimate the number of participants in 2011, as people who participated in continuing training several times were counted several times, unlike in the following years. Nevertheless, for certain questions, it can be useful to combine the variables *wb011\** in *wb010\** entirely, for example if it is “only” decisive whether employees with unskilled jobs have participated in a continuing training programme. The BIBB-FDZ has not programmed such a combination; the decision is up to each user. The corresponding commands for combining the variables are:

#### Stata

```
***Merging of variable wb011* (participants or cases of
participation of employees in continuing training) from 2011
cross-sectional data set
use "\Qualifizierungspanel_11bis17_long_gwa.dta"
sort id2 year
merge 1:n id2 year using "\Qualifizierungspanel_11bis17_
cross_2011_gwa.dta", keepusing(wb011inf* wb011qual* wb011high*)
drop _merge
***Replacing the variables
foreach x in einf qual high{
replace wb010`x'=wb011`x' if year==2011
replace wb010`x'_imp=wb011`x'_imp if year==2011
}
drop wb011*
```

#### SPSS

```
get file = "\Qualifizierungspanel_11bis17_long_gwa.sav".
Data set NAME long.
SORT CASES by id2 year.

get file = "\Qualifizierungspanel_11bis17_quer_2011_gwa.sav".
save OUTFILE= "\Qualifizierungspanel_11bis17_quer_2011_wb011.
sav"
/keep =id2 year wb011einf wb011einf_imp wb011qual wb011qual_imp
wb011hoch wb011hoch_imp.
get file= "\Qualifizierungspanel_11bis17_quer_2011_wb011.sav".
Data set NAME wb011.
SORT CASES by id2 year.
Data set ACTIVATE long.

MATCH FILES /FILE=*
/TABLE='wb011'
/by id2 year.
EXECUTE.

***Replacing the variables.
if year=2011 wb010einf=wb011einf
if year=2011 wb010qual=wb011qual.
if year=2011 wb010high=wb011high.
```

```

if year=2011 wb010einf_imp=wb011einf_imp.
if year=2011 wb010qual_imp=wb011qual_imp.
if jahr=2011 wb010hoch_imp=wb011hoch_imp.
EXECUTE.
DELETE VARIABLES wb011einf wb011einf_imp wb011qual wb011qual_imp
wb011hoch wb011hoch_imp.
EXECUTE.

```

#### 4.5.2 2012

The question of the occupation in which the establishment provided VET on 31 December of the corresponding year (*ab004\_a\*\_kldb2010\_5er*) and the question of the number of apprentices in the corresponding occupation (*ab004\_b1*) was only put to non-panel establishments in 2012 (*b44a\** and *b44b\**). The variable *ab004\_a\*\_kldb2010\_5er* in the long data set contains this information. 727 panel establishments that were not asked the question B44 but were VET establishments are marked with the value '99996. 2012 only new establishments queried'. In order to supplement the variables *ab004\_a\*\_kldb2010\_5er* for the year 2012, it is possible to use the information from the previous survey year (2011) and the information on the VET occupations of newly hired apprentices in 2012 (*b11a\**). However, the BIBB-FDZ has not included this recoding in the long data set. The decision about the recoding is up to each user; one way is:

##### Stata

```

***Merging of variable b11a* (VET occupation according to BBiG/
HwO 1 Kldb2010) from 2012 cross-sectional data set
use "\Qualifizierungspanel_11bis17_long_gwa.dta"
sort id2 year
merge 1:n id2 year using "\Qualification_panel_11to17_quer_2012_
gwa.dta", keepusing(b11a*)
drop _merge

***Take over VET occupations from the previous year if newly
filled VET occupations correspond to the VET occupations from the
previous year
foreach var of varlist ab004_a*_kldb2010_5er{
  forvalues x=1/5{
    replace `var'=b11a'x' if `var'==99996 & `var'[_n-1]==b11a'x' &
id2==id2[_n] & b11a'x'>0
  }
}
*if only 99999. No further VET occupation has been changed, back
to 99996. 2012 only new establishments surveyed code
recode ab004_a*_kldb2010_5er (99999=99996) if ab004_a1_
kldb2010_5er>9999995 & ab004_a2_kldb2010_5er>99995 & ab004_a3_
kldb2010_5er>99995 & ab004_a4_kldb2010_5er>99995 & ab004_a5_
kldb2010_5er>9999995

*if one VET occupation has been replaced, coding the others in
99999. No further VET occupation instead of 99996. 2012 only new

```

```

establishments surveyed
recode ab004_a*_kldb2010_5er (99996=99999) if ab004_a1_
kldb2010_5er<99996

*Swap information on VET occupations so that valid values are
first and then 99999. No further VET occupation
//swapval-ado must be downloaded; the command swaps the values
of the specified variables
forvalues x=3/5{
if ab004_a2_kldb2010_5er==99999{
cap swapval ab004_a`x`_kldb2010_5er ab004_a2_kldb2010_5er if
ab004_a`x`_kldb2010_5er<99996
}
}
forvalues x=4/5{
if ab004_a3_kldb2010_5er==99999{
cap swapval ab004_a`x`_kldb2010_5er ab004_a3_kldb2010_5er if
ab004_a`x`_kldb2010_5er<99996
}
}
if ab004_a4_kldb2010_5er==99999{
cap swapval ab004_a5_kldb2010_5er ab004_a4_kldb2010_5er if
ab004_a5_kldb2010_5er<99996
}
drop b11a*

```

### SPSS

```

get file ="*\Qualifizierungspanel_11bis17_long_gwa.sav".
Data set NAME long.
SORT CASES by id2 year.

GET file = ,*\Qualifizierungspanel_11bis17_quer_2012_gwa.sav".
SORT CASES by id2 year.
save OUTFILE= ,*\Qualifizierungspanel_11bis17_quer_2012_b11a.sav"
/KEEP= id2 year b11a1 b11a2 b11a3 b11a4 b11a5.
get file= ,*\Qualifizierungspanel_11bis17_quer_2012_b11a.sav".
Data set NAME b11a.
EXECUTE.
Data set ACTIVATE long.

MATCH FILES /FILE=*
/TABLE='b11a'
/BY id2 year.
EXECUTE.

***Take over VET occupations from the previous year if newly
filled VET occupations correspond to the VET occupations from the
previous year.
do repeat ab004_a = ab004_a1_kldb2010_5er ab004_a2_kldb2010_5er
ab004_a3_kldb2010_5er ab004_a4_kldb2010_5er ab004_a5_

```



```

kldb2010_5er.
  if ab004_a=99996 & LAG(ab004_a)=b11a1 & id2=LAG(id2) & b11a1>0
ab004_a=b11a1.
  if ab004_a=99996 & LAG(ab004_a)=b11a2 & id2=LAG(id2) & b11a2>0
ab004_a=b11a2.
  if ab004_a=99996 & LAG(ab004_a)=b11a3 & id2=LAG(id2) & b11a3>0
ab004_a=b11a3.
  if ab004_a=99996 & LAG(ab004_a)=b11a4 & id2=LAG(id2) & b11a4>0
ab004_a=b11a4.
  if ab004_a=99996 & LAG(ab004_a)=b11a5 & id2=LAG(id2) & b11a5>0
ab004_a=b11a5.
EXECUTE.

```

The additional data on newly hired (*b11\**) and unfilled (*b16\**) VET occupations in 2012 were originally collected as Kldb 1992 codes. In order to ensure comparability with the other occupational codes of the BIBB Training Panel, they were recoded in the Kldb 2010. This means that the occupation codes in the 2012 cross-sectional data set belonging to the long data set differ from the codes in the original cross-sectional data set.

The questions on *pb017\** (“Number of employees leaving the establishment by employee group”) were not asked in 2012. However, the question is the basis for the coding of the category ‘group of employees does not exist’. For this reason, the coding for 2012 must be made without the information on the number of employees leaving the establishment.

### 4.5.3 2011 and 2015

In 2011 and 2015, the survey asked which goals were associated with the continuing training measures of the previous year. Three of the items are comparable between the years (cf. Table 6). As only three out of seven items could be harmonised, they were not included in the long2 data set. Harmonisation is left to the respective users.

**Table 6: Comparable items on the objectives of continuing training measures in 2011 and 2015**

<i>Variable name 2011</i>	<i>Variable name 2015</i>	<i>Label</i>
a47b	e44_3	Aims of the continuing training: adaptation of technical knowledge to changing work processes
a47c	e44_7	Aims of the continuing training: to encourage employees to acquire knowledge independently
a47e	e44_2	Objectives of the continuing training: preparation for internal promotions

### 4.5.4 2016 and 2017

As already described in Chapter 2, a CATI survey focusing on technology and automation was conducted in 2016 in addition to the usual survey. Some of the variables were collected in both the CAPI and CATI surveys. Accordingly, some of the variables of the CATI and CAPI surveys are identical: they are documented in FRIEDRICH/GERHARDS 2019.

In addition to information on whether it is a cross-sectional case, the *quer* variable (cross-sectional case) also indicates 3,521 establishments that participated in the CATI survey

in 2016. Not all variables in the long and long2 data set are filled for these establishments. The corresponding cases have the value ‘-12 CATI case’ for all variables that do not belong to the question programme of the CATI survey.

Around 38 per cent (1,345) of the CATI establishments from 2016 participated in the panel again in 2017. For most panel analyses, it makes sense to exclude all exclusively CATI establishments. The corresponding commands are:

```
Stata
drop if quer==2 & pan_ges==1

SPSS
compute filter=1.
if quer=2 & pan_ges=1 filter=2.
EXECUTE.

select if filter=1.
EXECUTE.
DELETE VARIABLES filter.
EXECUTE.
```

Whether it makes sense to also exclude the CATI cases that took part in the BIBB Training Panel again in 2017 depends on the precise analytical interest. Some variables, such as the number of employees, are also filled for the CATI establishments and can therefore be evaluated in a longitudinal analyses.

## 4.6 English translation

The BIBB-FDZ also labelled the three longitudinal data sets of the BIBB Training Panel 2011 to 2017 in English and directly implemented the labels in Stata. The following command adjusts the language of the labels:

```
Stata
***Shows available languages
label language
***Changes the language to English
label language en
***Changes the language to German
label language de
```

For SPSS, the BIBB-FDZ provides three syntaxes, which re-label the variables and categories. These files are available in the BIBB-FDZ metadata portal and on the English BIBB-FDZ website.

## 5 Weighting and extrapolation

### 5.1 Calculation of weighting and extrapolation factors

With regard to weighting, it should be noted that in 2015, as was already described, the survey institute changed, which in turn led to a change in the way the weights are calculated. The exact description of the weighting can be found in the respective field reports of the survey institutions, e.g. HÄRING et al. 2017 and TNS INFRATEST 2015. A brief description of the weighting from 2015 onwards, following HÄRING et al. 2017, is given below.

The cross-sectional weights are a combination of (adjusted) design and structural adjustment weights. In general, the term “design weighting” is used when the selection probabilities resulting from the selection procedure are taken into account in the estimation in the form of weights as the inverse of the selection probability (cf. GABLER/GANNINGER 2010, p. 147). Non-response can result in a distortion in the distribution of certain characteristics in the sample. Adjustment weights should help to compensate imbalance due to non-response (cf. GABLER/GANNINGER 2010, p. 153). Through (structural) adjustment, the marginal distributions of the sample are adapted to the marginal distributions of the population (calibration).

The design weights are calculated separately for the refresher and panel sample. 128 stratification cells result from the stratification characteristics region, size, industry and VET (cf. Chapter 2). Using a Horvitz-Thompson estimator (cf. GABLER/GANNINGER 2010), the inverse selection probabilities are calculated for all strata and levels, and their reciprocals constitute the design weight. In order to avoid extreme weights, the design weights were normalised in terms of the number of cases.

For the panel sample, the calibrated weight from the last survey wave or, for temporary failures, from the penultimate one is defined as the design weight. Both design weights are adjusted separately by regression. The non-response adjustment of the design weights is again performed separately for refresher and panel establishments. For this purpose, default models are calculated whose independent variables represent the structural characteristics. The calculated participation probability is multiplied by the design weight to calculate the adjusted design weight.

The failure model for panel establishments also estimates the retention probability (*pbleib*) needed to calculate the longitudinal weighting. The longitudinal section weight (*dw\_p*) is estimated in the same way as the design weight, except that instead of the selection probability the retention probability is used for the calculation. In addition, the extrapolation factor (*hr\_p*) was calculated from the longitudinal section weight.

Additionally, the adjusted design weights of the refreshers and panel establishments are combined (*adjdw\_ges*). For this purpose, the adjusted design weight is multiplied by the quotient of the number of cases in the respective subsample (refreshers or panel establishments) and the total sample size.

Then the adjusted total design weight is calibrated by means of IPF (Iterative Proportional Fitting) of the combined stratification characteristics. The result is the structural adjustment weight (*sw\_q*) and the extrapolation factor (*hr\_q*). The calibration was additionally performed for industries (*hr\_br* and *sw\_br*) and employees (*hr\_bes* and *sw\_bs*).

## 5.2 Weighting and extrapolation factors in the data set

Most of the cross-sectional and longitudinal weights, extrapolation factors as well as the retention probability are contained in a separate data set “Qualifizierungspanel\_11bis17\_weight\_gwa”, whereas the variables *hr\_q* and *hr\_p* are contained in the long main data set (see Table 7). The data set including the weights is also stored in the long format; the weighting and extrapolation variables contain the values from the year of the survey.

The additional CATI survey in 2016 also affects the weighting. Separate weights were calculated for CAPI and CATI establishments and an integrated weight was calculated. In the data set, the variables *\*\_q*, *\*\_bes*, *\*\_br*, *\*\_p* and *adjdw\_ges* each contain the integrated weight. The individual weights are also contained in the data set “Qualifizierungspanel\_11bis17\_weight\_gwa”.

Table 7: Overview of weights or extrapolation factors contained in the data set

Variable	Data set	Variable content	Function	Years	Application	Number of cases	Mean value	Minimum	Maximum
sw_q	weight	Structural adjustment weight for the cross section	Adjusted design weighting and compensation for the distortion in the distribution of the stratification characteristics (i.e. industry, size, region and VET yes/no) for the population of establishments	Since 2015	<u>Analyses at establishment level for individual years</u> For example: what percentage of establishments in Germany offered continuing training	14,416	1	0.001	14.16
sw_bes	weight	Structural adjustment weight in proportion to employment	Adjusted design weighting and compensation of the distortion in the distribution of employees, in the population	Since 2015	<u>Analysis at employee level for individual years</u> For example: what percentage of the employees participate in continuing training	14,416	1	0.001	11.6
sw_br	weight	Structural adjustment weight for 20 industries	Adjusted design weighting and compensation of the distortion in the distribution of the industry, in the population	Since 2015	<u>Analyses at industry level for individual years</u> For example: the percentage of continuing training in manufacturing	14,416	1	0.001	24.7
hr_q	long	Extrapolation factor for the cross section of the respective year	Adjusted design weighting and compensation for distortion in the distribution of the stratification characteristics (i.e. industry, size, region and VET yes/no) and extrapolation to the population of establishments	2011-2017	<u>Analyses at establishment level for individual years</u> For example: how many establishments in Germany offered continuing training	23,902	607.6	0.39	11,984.1
hr_bes	weight	Extrapolation factor proportional to the number of employees	Adjusted design weighting and compensation for the distortion in the distribution of employees, in the population and extrapolation to the employees in the population	Since 2015	<u>Analysis at employee level for individual years</u> For example: how many employees participate in continuing training	14,416	398.6	0.3	6,061.1

Variable	Data set	Variable content	Function	Years	Application	Number of cases	Mean value	Minimum	Maximum
hr_br	weight	Extrapolation factor proportional for 20 industries	Adjusted design weighting and compensation of distortion in the distribution of the industry, in the population and extrapolation to the industries, in the population	Since 2013	<u>Analyses at industry level for individual years</u> For example: how much continuing training is offered in the construction industry	19,892	523.8	0.6	3,2167.2
dw_p	weight	Longitudinal design weight of the respective years	Adjusted design weighting and compensation for panel mortality and unexplained panel attendance	Since 2015	<u>Analyses over several years</u> For example: how did the percentage of establishments providing continuing training changed between 2011 and 2017?	6,935	1	0.001	23.5
hr_p	long	Extrapolation factor for longitudinal section of the respective years	Adjusted design weighting and compensation for panel mortality and unexplained panel attendance as well as extrapolation to the population of establishments	2012-2017	<u>Analyses over several years</u> For example: how did the number of establishments offering continuing training changed between 2011 and 2017?	10,472	1,187.8	0.7	32,167.2
adjdw_ges	weight	Modified design weight integrated total sample	Design weighting and non-response adjustment	Since 2015 <sup>9</sup>	Analyses without calibration of the stratification characteristics	10,735	1	0.001	32.6
pbleib	weight	Retention probability		Since 2015		2,021	0.6	0.1	0.8
<b>Special variables for individual years</b>									
hr11_13p hr11_14p hr12_14p	weight	Extrapolation factor for longitudinal section of the respective years		2013, 2014	Analyses for the years 2001 to 2013; or 2011 to 2014; or 2012 to 2014	860 553 786	-	-	-
*_capi *_cati	weight	Extrapolation factors, structural weightings, design weightings (adjusted and longitudinal)		2016	Analyses only for CATI or CAPI establishments	3,601 3,521	-	-	-

9 There is no corresponding weight for 2017 as there was no refresher sample.

For descriptive calculations of changes from one year to the next, a longitudinal section weight should be used. Each of the longitudinal section weights or extrapolation factors included in the data contain the factors to compensate for the retention probability from the previous year (or from two years earlier if the establishment did not participate in the previous year) to the survey year. When using longitudinal weights, it must be noted that no longitudinal section weighting or longitudinal extrapolation factor is available for the year of the initial interview. The year of the initial interview should be weighted or extrapolated with the respective cross-sectional factor. The commands (after the weight data set has been merged to the long data set) for combining the weights are:

**Stata**

```
replace hr_p= hr_q if well==1 | well==-3  
replace dw_p= sw_q if well==1 | well==-3
```

**SPSS**

```
if well==1 or well==-3 hr_p= hr_q  
if well==1 or well==-3 dw_p= sw_q  
EXECUTE.
```

## 6 Continuation in the following years

The BIBB-FDZ will continue to publish future waves of the BIBB Training Panel as an annual cross-sectional data set. In addition, the research data will be integrated into the existing longitudinal data set. This procedure implies the update and publication of the longitudinal data set of the BIBB Training Panel every year. There will be no separate documentation or updating of the data and methodological report on the longitudinal data set.



## 7 Anonymisation and data access

### 7.1 Data protection provisions

Access to the research data of the BIBB-FDZ is strictly subject to the principle of data economy and is compliant with the applicable data protection regulations in accordance with Regulation (EU) 2016/679 (Basic Data Protection Regulation, GDPR) and supplementary provisions of the German Data Protection Act (BDSG 2018). Accordingly, data can be passed on to independent scientific research for carrying out scientific projects if it is not possible to establish a reference to a survey unit (“anonymity”). It is recommended that researchers who are authorised to access and use the data be particularly obliged to observe data protection regulations (cf. BIBB-FDZ Guidelines of the Research Data Centre at BIBB for Remote Data Access).

BIBB-FDZ staff only gain insight into research questions, methods and analyses conducted by the researchers for the purposes of providing advice, improving the BIBB-FDZ service and ensuring compliance with data protection. BIBB employees who do not belong to the BIBB-FDZ do not gain insight into the activities of the researchers.

### 7.2 Formal data anonymisation

Prior to the survey, the voluntarily participating establishments were assured data protection and anonymity (see data protection sheet in HÄRING et al. 2017). The little structural information in the data set does not allow for a re-identification of the surveyed establishments. To differentiate between the cases, a system-independent establishment code applicable across all waves (*id2*) is available.

### 7.3 Data access

The data from the BIBB Training Panel 2011 to 2017 can be analysed via remote data access (RDA) and the on-site visit at BIBB in Bonn.

Remote data access processing allows the analysis of weakly anonymised research data via the processing of executable user-created syntax programmes. Data users can download structural data sets (test data) from the BIBB-FDZ metadata portal or English website. The structure and characteristics of the test data are similar to the original data. They thus enable the creation programme codes (in Stata or SPSS) which the BIBB-FDZ then uses to analyse the original data. The output, which has been checked for data confidentiality, is then sent back to the users. On-site visits allow the analysis of weakly anonymised research data at the isolated PC workstations in a separate guest room at the BIBB-FDZ.

In 2021, the BIBB-FDZ will put into operation an automated remote data processing system, which will enable data users to apply their own syntax programmes to original data. Evaluation results can then be viewed (personalised) immediately after the evaluation programmes have been run. In addition, the two data access paths, remote data access and on-site visits will be linked in this way on a project-related basis, i.e. access to the same and current project directory will be possible on both data access paths. The introduction of the automated remote data access system will completely replace the manual procedure described in the previous paragraph for the current remote data access at the BIBB-FDZ. This will give data users greater

autonomy and a higher degree of freedom in the evaluation of sensitive data sets, which cannot be made available via a SUF due to possible re-identification risks.

The use of the data requires a formal application for all data access routes described here. The corresponding applications can be downloaded from the BIBB-FDZ website.

## 8 References

- FISCHER, Gabriele u.a.: Das IAB-Betriebspanel – von der Stichprobe über die Erhebung bis zur Hochrechnung. FDZ Methodenreport 01/2008, URL: [http://doku.iab.de/fdz/reporte/2008/MR\\_01-08.pdf](http://doku.iab.de/fdz/reporte/2008/MR_01-08.pdf) (Access: 10.08.2020)
- FRIEDRICH, Anett; GERHARDS, Christian: BIBB-Qualifizierungspanel 2015. Daten- und Methodenbericht Nr. 2/2017. Bonn 2017
- FRIEDRICH, Anett; GERHARDS, Christian: BIBB-Qualifizierungspanel 2016. Daten- und Methodenbericht Nr. 1/2018. Bonn 2018
- GABLER, Siegfried; GANNINGER, Matthias: Gewichtung. In: WOLF, Christof; BEST, Henning (Eds.): Handbuch der sozialwissenschaftlichen Datenanalyse. Wiesbaden 2010, p. 143–164
- GERHARDS, Christian; FRIEDRICH, Anett: BIBB-Qualifizierungspanel 2014. Daten- und Methodenbericht Nr. 3/2016. Bonn 2016
- GERHARDS, Christian u.a.: The BIBB Training Panel. An Establishment Panel on Training and Competence Development. Schmollers Jahrbuch (2012) 132(4), p. 635-652
- GERHARDS, Christian u.a.: BIBB-Qualifizierungspanel 2013. Daten- und Methodenbericht Nr. 3/2015. Bonn 2015
- HÄRING, Armando; SCHIEL, Stefan; KLEUDGEN, Martin: Methodenbericht Erhebungswelle 2017 des BIBB-Qualifizierungspanels. Bonn 2017
- HOHN, Christine; FRIEDRICH, Anett; GERHARDS, Christian: Querschnitt- und Panelcharakteristika des BIBB-Qualifizierungspanels. Bonn 2020. URL: <https://metadaten.bibb.de/metadatengruppe/16>. (Access: 12.10.2020)
- TNS INFRATEST SOZIALFORSCHUNG: BIBB-Qualifizierungspanel. Qualifizierung und Kompetenzentwicklung. Methodenbericht zur vierten Welle. München 2015

# Appendix

Table A1: Variables not queried in the "Qualifizierungspanel\_11bis17\_long\_gwa" in all years

Variable(s)	Label	Not queried years
ab023_*	School-leaving certificate applicants	2011, 2013
ab034*	Number of vacant VET positions because VET contracts were terminated prematurely	2011, 2012
ab036	VET contracts in the past calendar year terminated prematurely that year	2012, 2013
ab037	Number of VET contracts in the past calendar year terminated prematurely that year	2012, 2013
ab057	Trained apprentices or offered VET according to BBiG/HwO in the last 3 years	2011 to 2014
ab065	Any refugees among the apprentices in your establishment on 31.12. last year (according to BBiG/HwO)	2011 to 2015
ab066	Number of refugees among apprentices according to BBiG/HwO on 31.12. last year	2011 to 2015
ab067	Number of refugees trained in occupations not regulated according to BBiG/HwO on 31.12. last year	2011 to 2015
ab068	Number of newly hired refugees trained according to BBiG/HwO the current VET year	2011 to 2015
ab069	VET contracts according to BBiG/HwO with refugees cancelled before 31.12.	2011 to 2015
ps001zwi*	Subtotal all employees without apprentices	2011
ps001aus*	Apprentices according to BBiG/HwO	2011
ps001soz*	Apprentices according to VET regulations of healthcare or social/pedagogical professions	2011
ps001bea*	Prospective civil servants	2011 to 2014
ps001svb_v	Previous year: employees subject to social insurance contributions without apprentices excluding marginal employees	2011
ps001ger_v	Previous year: marginal employees, 400 or 450 € employees, short-term employees	2011
ps001nsv_v	Previous year: active owners/assisting family members, other employees not subject to social insurance contributions	2011
ps030_sch	Employees with school-based VET degree, e.g. at vocational/technical college	2011, 2012
ps036	Were there refugees among the employees in your establishment on 31.12. last year?	2011 to 2014
ps037	Number of refugees among employees on 31.12. last year	2011 to 2014
pb001_ang	Did your establishment offer jobs in the last calendar year?	2011
pb005	Number of newly hired employees by employee group	2012
pb011_*	Number of vacant positions by employee group	2012
pb017	Number of employees who left the establishment by employee group	2012

Variable(s)	Label	Not queried years
wb002	Number of employees who participated in advanced VET in the last calendar year	2011, 2012
wb010*	Number of continuing training participants by employee group	2011, 2013
di001_*	Used technology	2011 to 2015
be001_*	Regulations	2011, 2013
be010*	Changes in the establishment	2011
<b><i>Items that were not consistently queried</i></b>		
ab043_*1	Retained permanently: number of apprentices	2012
ab043_*2	Retained temporarily: number of apprentices	2012
ab043_*4	Not retained for operational reasons: number of apprentices	2012 to 2014
ab043_*5	Not retained because of the examination marks: number of apprentices	2012 to 2014
ab043_*6	Not retained because not passed the exams: number of apprentices	2017
ab058_5	Reasons for VET: tradition of our establishment	2011
ab058_6	Reasons for VET: train skilled workers who can be employed in our establishment after their VET	2011 to 2015
ab058_7	Reasons for VET: VET as a joint task of industry	2011 to 2015
be001_6	Regulations: long-term accounts for working hours	2012, 2014
be001_7	Regulations: changing composition of project teams	2012, 2014
be001_8	Regulations: measures to reconcile work and family life	2012, 2014
be001_9	Regulations: old age provision	2012, 2014
be010_7	Changes: closure, outsourcing or spin-off of parts of the business	2012, 2013
be010_8	Changes: integration of other establishments or parts of other establishments	2012, 2013

Table A2: List of variables for the data set "Qualifizierungspanel\_11bis17\_long2\_gwa"

Variable	Variable label	Years surveyed
id2	System-independent identification number	
jahr	Year of the survey	
well	Wave identification	
quer	Cross-sectional case	
pan_ges	Number of participations until 2017	
ab003f	Number of female apprentices according to BBiG or Hw0 as of 31.12. of the previous year	2011-2015
ab003f_imp	Number of female apprentices as of 31.12. of the previous year	2011, 2012, 2014
ab003gt	Number of industrial-technical apprentices as of 31.12. of the previous year	2011-2015
ab003kv	Number of commercial-administrative apprentices as of 31.12. of the previous year	2011-2015
ab003gt_imp	Number of industrial-technical apprentices as of 31.12. of the previous year (imputed)	2012
ab003kv_imp	Number of commercial-administrative apprentices as of 31.12. of the previous year (imputed)	2012
ab009	Establishment had provided VET in more than 5 occupations	2016, 2017
ab027_1	Offers for Apprentices: VET remuneration above the union wages	2015, 2016
ab027_2	Offers for Apprentices: possibility to acquire certified additional qualifications	2015
ab027_3	Offers for Apprentices: takeover guarantee upon successful completion of VET	2015, 2016
ab027_4	Offers for Apprentices: monetary benefits (e.g. company car)	2015, 2016
ab027_5	Offers for Apprentices: additional programmes to support underachieving young people	2015
ab027_6	Offers for Apprentices: subsidy for the use of local transport / meals / accommodation	2015
ab027_7	Offers for Apprentices: subsidy for the use of public transport	2016
ab027_8	Offers for Apprentices: subsidy for board or accommodation	2016
ab035	Apprentices employed according to BBiG/Hw0 in the calendar year	2015-2017
ab055	Number of apprentices taken on for a limited period, who are expected to receive a permanent contract	2015, 2016
ab055_imp	Number of apprentices taken on for a limited period, who are expected to receive a permanent contract (imputed)	2015, 2016
ab056	VET positions planned according to BBiG/Hw0	2015-2017
ab064	Does your establishment fulfil the legal requirements for VET according to BBiG/ Hw0?	2015-2017
ab082gt1	Number of industrial-technical apprentices in the 1. year of VET	2011-2013
ab082gt2	Number of industrial-technical apprentices in the 2. year of VET	2011-2013

Variable	Variable label	Years surveyed
ab082gt3	Number of industrial-technical apprentices in the 3. year of VET	2011-2013
ab082gt4	Number of industrial-technical apprentices in the 4. year of VET	2011-2013
ab082kv1	Number of commercial-administrative apprentices in the 1. year of VET	2011-2013
ab082kv2	Number of commercial-administrative apprentices in the 2. year of VET	2011-2013
ab082kv3	Number of commercial-administrative apprentices in the 3. year of VET	2011-2013
ab082kv4	Number of commercial-administrative apprentices in the 4. year of VET	2011-2013
ab082gt1_imp	Apprentices in the 1. year of VET industrial-technical (imputed)	2011
ab082gt2_imp	Apprentices in the 2. year of VET industrial-technical (imputed)	2011
ab082gt3_imp	Apprentices in the 3. year of VET industrial-technical (imputed)	2011
ab082gt4_imp	Apprentices in the 4. year of VET industrial-technical (imputed)	2011
ab082kv1_imp	Apprentices in the 1. year of VET commercial-administrative (imputed)	2011
ab082kv2_imp	Apprentices in the 2. year of VET commercial-administrative (imputed)	2011
ab082kv3_imp	Apprentices in the 3. year of VET commercial-administrative (imputed)	2011
ab082kv4_imp	Apprentices in the 4. year of VET commercial-administrative (imputed)	2011
ab083	Apprentices doing dual studies as of 31.12. of the previous year	2011, 2014
ab083gt	Number of industrial-technical apprentices doing dual studies as of 31.12. of the previous year	2011, 2014
ab083kv	Number of commercial-administrative apprentices doing dual studies as of 31.12. of the previous year	2011, 2014
ab083gt_imp	Number of industrial-technical apprentices doing dual studies as of 31.12. of the previous year	2014
ab083kv_imp	Number of commercial-administrative apprentices in dual studies as of 31.12. of the previous year	2014
ab085gt	Engagement in training seminars or courses, industrial-technical apprentices	2011-2013
ab085kv	Engagement in training seminars or courses, commercial-administrative apprentices	2011-2013
ab086gt1	Share of productive activities during presence in the establishment in the 1. year of VET, industrial-technical apprentices	2011-2013
ab086gt2	Share of productive activities during presence in the establishment in the 2. year of VET, industrial-technical apprentices	2011-2013
ab086gt3	Share of productive activities during presence in the establishment in the 3. year of VET, industrial-technical apprentices	2011-2013
ab086gt4	Share of productive activities during presence in the establishment in the 4. year of VET, industrial-technical apprentices	2011-2013
ab086kv1	Share of productive activities during presence in the establishment in the 1. year of VET, commercial-administrative apprentices	2011-2013
ab086kv2	Share of productive activities during presence in the establishment in the 2. year of VET, commercial-administrative apprentices	2011-2013
ab086kv3	Share of productive activities during presence in the establishment in the 3. year of VET, commercial-administrative apprentices	2011-2013

Variable	Variable label	Years surveyed
ab086kv4	Share of productive activities during presence in the establishment in the 4. year of VET, commercial-administrative apprentices	2011-2013
ab087gt1	Perceptual performance level of industrial-technical apprentices in the 1. year of VET	2011-2013
ab087gt2	Perceptual performance level of industrial-technical apprentices in the 2. year of VET	2011-2013
ab087gt3	Perceptual performance level of industrial-technical apprentices in the 3. year of VET	2011-2013
ab087gt4	Perceptual performance level of industrial-technical apprentices in the 4. year of VET	2011-2013
ab087kv1	Perceptual performance level of commercial-administrative apprentices in the 1. year of VET	2011-2013
ab087kv2	Perceptual performance level of commercial-administrative apprentices in the 2. year of VET	2011-2013
ab087kv3	Perceptual performance level of commercial-administrative apprentices in the 3. year of VET	2011-2013
ab087kv4	Perceptual performance level of commercial-administrative apprentices in the 4. year of VET	2011-2013
ab089	Forecast: number of VET positions in the next 3 years	2011-2013
ps039	Number as of 31.12. of the previous year: employees with a bachelor's degree as the highest degree	2013, 2014, 2017
ps039_imp	Number as of 31.12. of the previous year: employees with a bachelor's degree as the highest degree (imputed)	2013, 2014
wb015_1	Promotions	2013, 2014
wb015_2	Number of employees promoted	2013, 2014
wb015_2_imp	Number of employees promoted (imputed)	2013, 2014
wb016_1	Change in the scope of tasks/activities	2013, 2014
wb016_2	Number of employees changing their tasks/activities	2013, 2014
wb016_2_imp	Number of employees changing their tasks/activities (imputed)	2013, 2014
wb017_einf1	Employees with unskilled tasks: participation in lectures, conferences	2012, 2014, 2015, 2017
wb017_einf2	Employees with unskilled tasks: participation in quality circles, learning groups	2012, 2014, 2015, 2017
wb017_einf3	Employees with unskilled tasks: instructions/training at the workplace	2012, 2014, 2015, 2017
wb017_einf4	Employees with unskilled tasks: mentoring programmes, individual counselling, coaching	2012, 2014, 2015, 2017
wb017_einf5	Employees with unskilled tasks: self-directed learning	2012, 2014, 2015, 2017
wb017_qual1	Employees with qualified tasks: participation in lectures, conferences	2012, 2014, 2015, 2017
wb017_qual2	Employees with qualified tasks: participation in quality circles, learning groups	2012, 2014, 2015, 2017
wb017_qual3	Employees with qualified tasks: instructions/training at the workplace	2012, 2014, 2015, 2017



Variable	Variable label	Years surveyed
wb017_qual4	Employees with qualified tasks: mentoring programmes, individual counselling, coaching	2012, 2014, 2015, 2017
wb017_qual5	Employees with qualified tasks: self-directed learning	2012, 2014, 2015, 2017
wb017_hoch1	Employees with highly qualified tasks: participation in lectures, conferences	2012, 2014, 2015, 2017
wb017_hoch2	Employees with highly qualified tasks: participation in quality circles, learning groups	2012, 2014, 2015, 2017
wb017_hoch3	Employees with highly qualified tasks: instructions/training at the workplace	2012, 2014, 2015, 2017
wb017_hoch4	Employees with highly qualified tasks: mentoring programmes, individual counselling, coaching	2012, 2014, 2015, 2017
wb017_hoch5	Employees with highly qualified tasks: self-directed learning	2012, 2014, 2015, 2017
wb018	Implementation of continuing training measures due to legal regulations	2013, 2014
di018_1	Effects of digital technologies: relieving employees of routine tasks	2016, 2017
di018_2	Effects of digital technologies: better opportunities for the low-skilled employees through digital support	2016, 2017
di018_3	Effects of digital technologies: elimination of certain tasks and occupations	2016, 2017
di018_4	Effects of digital technologies: highly qualified employees are relieved of routine tasks	2016, 2017
di018_5	Effects of digital technologies: demand for skilled workers will decrease	2016, 2017
di018_6	Effects of digital technologies: jobs will be lost	2016, 2017
mo001einf_1	Employees with unskilled tasks: frequency: tasks are specified in every detail	2012, 2015, 2016
mo001einf_2	Employees with unskilled tasks: frequency: tasks are repetitive	2012, 2015, 2016
mo001einf_3	Employees with unskilled tasks: frequency: using tools/machines	2012, 2015, 2016
mo001einf_4	Employees with unskilled tasks: frequency: dexterity/manual skills	2012, 2015, 2016
mo001einf_5	Employees with unskilled tasks: frequency: consulting customers or patients	2012, 2015, 2016
mo001einf_6	Employees with unskilled tasks: frequency: convincing others/negotiate compromises	2012, 2015, 2016
mo001einf_7	Employees with unskilled tasks: frequency: organising or researching	2012, 2015, 2016
mo001einf_8	Employees with unskilled tasks: frequency: improving/re-testing procedures and processes	2012, 2015, 2016
mo001qual_1	Employees with qualified tasks: frequency: tasks are specified in every detail	2012, 2015, 2016
mo001qual_2	Employees with qualified tasks: frequency: tasks are repetitive	2012, 2015, 2016
mo001qual_3	Employees with qualified tasks: frequency: using tools/machines	2012, 2015, 2016
mo001qual_4	Employees with qualified tasks: frequency: dexterity/manual skills	2012, 2015, 2016

Variable	Variable label	Years surveyed
mo001qual_5	Employees with qualified tasks: frequency: consulting customers or patients	2012, 2015, 2016
mo001qual_6	Employees with qualified tasks: frequency: convincing others/negotiate compromises	2012, 2015, 2016
mo001qual_7	Employees with qualified tasks: frequency: organising or researching	2012, 2015, 2016
mo001qual_8	Employees with qualified tasks: frequency: improving/re-testing procedures and processes	2012, 2015, 2016
mo001hoch_1	Employees with highly qualified tasks: frequency: tasks are specified in every detail	2012, 2015, 2016
mo001hoch_2	Employees with highly qualified tasks: frequency: tasks are repetitive	2012, 2015, 2016
mo001hoch_3	Employees with highly qualified tasks: frequency: using tools/machines	2012, 2015, 2016
mo001hoch_4	Employees with highly qualified tasks: frequency: dexterity/manual skills	2012, 2015, 2016
mo001hoch_5	Employees with highly qualified tasks: frequency: consulting customers or patients	2012, 2015, 2016
mo001hoch_6	Employees with highly qualified tasks: frequency: convincing others/negotiate compromises	2012, 2015, 2016
mo001hoch_7	Employees with highly qualified tasks: frequency: organising or researching	2012, 2015, 2016
mo001hoch_8	Employees with highly qualified tasks: frequency: improving/re-testing procedures and processes	2012, 2015, 2016
be022	Development of business volume compared to the previous year	2015-2017
be042_1	Importance as a competitive factor: high-quality products/services	2011, 2013
be042_2	Importance as a competitive factor: new types of products and services	2011, 2013
be042_3	Importance as competitive factor: low price	2011, 2013

Table A3: Harmonisation applied to all variables and indications for comparison over time

Variable(s)	Label	Year(s)	Differences	Procedure
<b>BIBB Qualifizierungspanel_11bis17_long_gwa</b>				
<b>Topic: VET</b>				
ab001	Did your establishment employ apprentices as of 31.12. last year?	2012 to 2014	The question was asked in combination with ab002.	The variable has been coded to the system from 2015 onwards. Categories 1 to 3 have been combined into '1=yes'.
ab002	Exclusively apprentices according to BBIG/HwO or also according to other regulations	2011	The question was not asked.	The cases have been coded accordingly.
		2012 to 2014	The question was asked in combination with ab001, and the wording of the answer categories differs from the subsequent years.	The information from ab001 was transferred.
ab004a*kldb-2010 (BBIG/HwO) as of 31.12. last year	VET occupation Kldb 2010 (BBIG/HwO) as of 31.12. last year	2011 to 2013	The variable includes Kldb 1992 codes.	All variables were recoded to Kldb 2010. The Kldb 1992 is not included in the data set.
		2011	In this year, the formulation of the questions differs from the other years: no explicit questions were asked about apprentices in BBIG/HwO occupations.	
		all	In the cross section, the category 'No further VET occupation' is not coded uniformly.	In all waves the value '99999. No further VET occupation' was coded.
ab004*	Classification Production/Service occupation VET occupation	all	The variables were coded partly from Kldb 1992 and partly from Kldb 2010.	All variables were coded from the Kldb 2010 contained in the data set uniformly.
ab004_b*	Number of apprentices as of 31.12. last year VET occupation	all	In the cross sections, the category 'No further VET occupations' is not coded uniformly.	In all waves the value '99999. No further VET occupation' was coded.

Variable(s)	Label	Year(s)	Differences	Procedure
ab015	Hired new apprentices in the current VET year according to BBig/Hw0	2011	In 2011, the question was not asked in the same way as in the subsequent years.	Question a29a was used to generate the information for the variable. All cases where a29a==1, i.e. "VET places were offered" and "Apprentices were hired" are coded as ab015==1, i.e. yes, new apprentices hired. All cases where a29a==2, i.e. "VET places offered but no apprentices recruited" are coded as ab015==2, i.e. no new apprentices recruited. All cases where a29a==3, i.e. "No VET places" offered were coded as ab015==2, i.e. no new apprentices recruited; under the assumption that if no VET places were offered, no apprentices could be recruited.
		2011; 2012	The filter guidance deviates from the following years. In these years, the question was also put to establishments that stated that they did not have any apprentices according to BBig/Hw0 (ab001==1 or ab002==3). From 2013, these establishments are filtered directly to the next question.	
ab016	Offered VET positions for the current VET year according to BBig/Hw0	2011	Instead of the answer options "Yes" and "No", there are three options: 1: Yes, has offered and hired 2: Yes, offered and did not hire anyone 3: No	The first answer category was coded as a filter because in the subsequent years the question was only put to establishments that did not hire any new apprentices. The second answer category was coded as "Yes", and the third answer category as "No". Due to the inconsistency of the questions, it is possible that in 2011 more establishments were filtered than in the following years and the answer category '2 no' is underestimated. The original variable a29a remains in the respective cross-sectional data set.
ab017	Number of hired apprentices according to BBig/Hw0	2011	The values were queried separately for commercial-administrative and industrial-technical professions.	The values were added together.
ab018*	School-leaving certificate new apprentices	2011, 2012	In 2011 and 2012, it was asked about the school-leaving certificate of all apprentices. In the following years, it was asked about apprentices in BBig/Hw0 occupations only. Accordingly, the number tends to be higher in 2011 and 2012.	

Variable(s)	Label	Year(s)	Differences	Procedure
ab027*	Offers for apprentices	2015, 2016	The items queried are not uniform over the years.	The categories '9996. item 2016 not queried' and '9995. item 2015 not queried' have been coded. Theoretically, the variables ab027_7 and ab027_08 could be combined to the variable ab027_6, but this was not done in the data; it is left to the users.
ab032	Offered VET positions for the current VET year that could not be filled (according to BBIG/Hw0)	From 2013	The following note was added to the question: "Please also take into account the VET places for the VET year that are not filled due to terminated VET contracts".	
ab033	Number VET positions not filled (according to BBIG/Hw0)	2011	The values were queried separately for commercial-administrative and industrial-technical professions.	The values were added together.
		From 2013	The following note was added to the question: "Please also take into account the VET places for the VET year that are not filled due to terminated VET contracts".	
ab036	VET contracts in the past calendar year terminated prematurely that year	2011	The information does not refer to the calendar year as in the other years, but to the year of VET.	Category 3 has been coded as '-1 filter'.
		2014	The categories differ from the other years. There is an additional category '3 Not applicable, had in the year 2013 no apprentices according to BBIG or Hw0'.	

Variable(s)	Label	Year(s)	Differences	Procedure
ab037	Number of VET contracts in the past calendar year terminated prematurely that year	2011	<p>The information does not refer to the calendar year as in the other years, but to the year of VET.</p> <p>The query was carried out separately for industrial-technical and commercial-administrative professions.</p> <p>The question was formulated differently.</p>	<p>Both values were added together. In one case the combination '-9 no specification' for commercial-technical and zero contract solution for commercial-administrative results in the value 0.</p>
ab038	Did apprentices take part in final examinations in the previous calendar year (according to BBlG/Hw0)	2013	<p>From 2014 onwards, question E14 (contract terminations of VET contracts in the calendar year 2015) was filtered and thus passed on to the next topic of the questionnaires when establishments stated that they had no apprentices according to BBlG or Hw0. As a result, these establishments do not answer question E16 (whether final examinations for apprentices under BBlG or Hw0 took place) and have the value '-1 filter' for the corresponding variable, which means that the category '3 No apprentices' no longer applies. In 2013, on the other hand, the question about final examinations without a previous filter was asked.</p>	<p>The cases of this category 3 have been coded as '-1 filter'.</p>
ab041*	Number of apprentices: passed final examination	2012 to 2014	<p>The question has not been asked during these years.</p>	<p>The values were calculated as the difference between the number of apprentices who took the final examination (ab039*) and those who failed the examination (ab043*).</p> <p>The filter guidance has been adjusted so that all cases have the value '-1 filter' if they were also filtered for ab039*.</p>

Variable(s)	Label	Year(s)	Differences	Procedure
ab043*	Retention of apprentices	all	The item formulation varies between the years. The wording is the same in 2011, 2015 and 2016, and the same in 2013 and 2014 (but one additional item). The 3rd item is the same for all years. The other differences are documented in the following rows.	
ab043_1 and ab043_2	Retained permanently Retained temporarily	2015	Only in 2015, the category '9999. no industrial-technical (or commercial-administrative) apprentices' was coded.	The coding has been adapted by coding the category '9999. no industrial/technical (or commercial/administrative) apprentices' in '-1 filter'.
ab043_2	Retained temporarily	2012	The items were not queried. Instead, the following items were asked: "Were retained as a skilled worker" and "Were retained in another position".	Included in the respective cross-sectional data set as variables b19*a and b19*b.
ab043_4	Not retained for operational reasons	2013, 2014	The items were formulated differently.	
ab043_5	Not retained because of the examination marks	2012 to 2014	The items were not queried.	
ab043_6	Not retained because not passed the exams	2012 to 2013 2014	The items were not queried. The item was queried together with "Have passed the test and not retained for operational reasons".	Included in the respective cross-sectional data set as variable d17*d.
		2011, 2015, 2016	The items were not collected.	The values were calculated as the difference between the number of apprentices, the number of apprentices who took part in the examination (ab039*) and the number of examinations passed (ab041*).

Variable(s)	Label	Year(s)	Differences	Procedure
ab058_*	Reasons for VET	2011	The question was worded differently, and a different scale was used. The question was put to all establishments, not only those that provide VET. Every user should critically examine whether a time comparison makes sense. Furthermore, item_4 "VET causes high costs" was only surveyed in 2011 and item_5 "Tradition of the establishment" was not surveyed.	
		From 2015	There is the additional 5th category	For a time comparison, categories 4 and 5 can be combined; this was not done in the data set, but is left to the user.
		2016	The items _6 "Train skilled workers who will remain long-term" and _7 "VET as a joint task of the economy" are only available this year.	
ab087*	Perceptual performance level of apprentices	2011, 2012, 2013	In cross section, the category "VET year not filled" was not coded uniformly.	The category "999. VET year not filled" has been coded uniformly.
<b>Topic: Staff</b>				
ps020*	Number of employees by group of employees	2011	With regard to the number of employees, a distinction is made between four groups of employees (unskilled, qualified usually apprenticeship completed, qualified usually master craftsman or technician completed, highly qualified jobs) instead of three (unskilled, qualified, highly qualified jobs).	The categories "Qualified usually apprenticeship completed" and "Qualified usually master craftsman or technician completed" were combined to "Employees with qualified tasks".
pb005*	Number of newly hired employees	2011	Among the newly recruited employees, a distinction is made between four groups of employees (unskilled, qualified usually apprenticeship completed, qualified usually master craftsman or technician completed, highly qualified jobs) instead of three (unskilled, qualified, highly qualified jobs).	The categories "Qualified usually apprenticeship completed" and "Qualified usually master craftsman or technician completed" were combined to "Employees with qualified tasks".



Variable(s)	Label	Year(s)	Differences	Procedure
pb011*	Number of vacant positions in the last calendar year	2011	With regard to the number of unfilled vacancies, a distinction is made between four groups of employees (unskilled, qualified usually apprenticeship completed, qualified usually master craftsman or technician completed, highly qualified jobs) instead of three (unskilled, qualified, highly qualified jobs).	The categories "Qualified usually apprenticeship completed" and "Qualified usually master craftsman or technician completed" were combined to "Employees with qualified tasks".
pb017*	Number of employees who left the establishment	2017	Question clearly filtered, therefore programming 9999 not necessary.	The category '9999. no employees retired' was replaced by -1.
<b>Topic: Continuing education</b>				
wb008, wb009	Participations in other continuing training, like internal/external courses, seminars/ training courses in the last calendar year.  Number of employees taking part in continuing training	From 2015	From 2015, the question is presented with the question of employees' participation in advanced VET (wb001). It can be assumed that advanced training courses are not included in wb008. Prior to 2015, the question of advanced VET was downstream or has not yet been asked. It can be assumed that without any specific demand, advanced VET courses are included in wb008. In addition, only from 2015 onwards, the question on the number of participants in continuing education (wb009) will refer to the following 'Please do not include participants in advanced VET or in a part-time study programme'.  For a time comparison it may therefore be useful to combine the variables wb001 and wb008 or wb002 and wb009 (and e40b 'part-time study programme' for 2015). However, this decision is left to each user. There is no harmonisation by the FDZ.	
wb010*	Number of continuing training participants	2011	When asked about the scope of continuing training in 2011, the number of participants (persons) or cases of participation (wb011*) was used, whereas in the other years the number of continuing training participants (wb010*) was used.	Based on the information whether persons or participation cases were reported (a44), the information is moved from variable wb011* to variable wb010*. For the other cases the code '9995. participation cases reported' is assigned.

Variable(s)	Label	Year(s)	Differences	Procedure
wb017_1- wb017_5	Forms of continuing training	2012, 2014, 2015, 2017  2011	The items were queried separately for unskilled, qualified and highly qualified employees.  The coding of the possible answers is '0 no' instead of '2 no'.	The values of the three variables were combined. The original variables are contained in the long2 data set.  The coding is adjusted '0 no' becomes '2 no'.
<b>Topic: General information about the establishment</b>				
wz	Industry	2011	In 2011, 45 instead of 43 industry categories were surveyed.	The 2011 industry codes have been recoded (old=new) as follows: 39=41; 40=42; 41=43; 42=44; 43=45; 44=46; 38=47 (47 being a special code that exists only for 2011).  The WZ codes 38, 39 and 40 are not assigned for 2011.
bra	Industry (8 categories)	2011 to 2013	The allocation of industry to categories (which also serve for stratification) has been changed.	The years 2011-2013 have been recoded according to the current classification. The corresponding code is in FRIEDRICH/GERHARDS 2017.
ti_ki	Technological intensity/ knowledge intensive services	All		The variable was not taken from the cross-section, but was newly created for the longitudinal data set.
be001*	Regulations	2012, 2014	The items queried are not uniform over the years.	The categories '9996 Item 2012 and 2014 not queried' have been coded.
be010_*	Changes	2012, 2013	The information for 2012 and 2013 refers to changes in the last two years. From 2014 onwards, changes in the previous year were asked for. Furthermore, the question 2013 was only put to new establishments, i.e. not to panel establishments.	
be018_1	Type of establishment	2012, 2013	The items queried are not uniform over the years.	The categories '9996 Item 2012 and 2013 not queried' were coded.
		2012	The questions were only put to panel establishments in 2012.	For panel establishments, the information from the previous year was taken over.
		From 2015	The order of the queried categories has been changed.	The years before 2015 have been adjusted to the changed order.

Variable(s)	Label	Year(s)	Differences	Procedure
be018_2	Business volume	2012	Panel establishments were not asked directly about their business volume but whether it had declined or increased, and if so by what percentage.	The business volume in 2011 and the information on the change from 2012 were used to calculate the business volume (done by the project). The information provided by the panel establishments and the establishments first interviewed was combined.
be020	Share of your foreign turnover as a proportion of total turnover	2012	Panel establishments were not asked directly about the share of foreign sales but whether it has fallen or risen, and if so by what percentage.	Using the share of foreign sales in 2011 and the data on the change from 2012, the share of foreign sales in total sales was calculated (done by the project). The information provided by the panel establishments and the establishments first interviewed was combined.
be021	Share of preliminary work and external costs in total sales	2012	Panel establishments were not asked directly about the share of preliminary work but whether it has fallen or risen, and if so by what percentage.	The share of preliminary work in 2011 and the data on the change from 2012 were used to calculate the share of preliminary work and external costs in total sales (done by the project). The information provided by the panel establishments and the establishments first interviewed was combined.
be023	Share of turnover accounted for by preliminary work and external costs	2012	Panel establishments were not asked directly about the share of turnover but whether it had fallen or risen, and if so by what percentage.	With the help of the share of turnover in 2011 and the data on the change from 2012, the share of turnover in the establishment in the last financial year was calculated (done by the project). The information provided by the panel establishments and the establishments first interviewed was combined.

Variable(s)	Label	Year(s)	Differences	Procedure
		2012	Panel establishments were not asked directly about gross wages but whether they had fallen or risen, and if so by what percentage.	With the help of the gross wage 2011 and the data on the change from 2012, the average gross wage was calculated (done by the project). The information provided by the panel establishments and the establishments first interviewed was combined.
be026*	Average gross wage	All	The category 'Employee group not exists' was coded differently in the survey years.	The category '999999. Employee group not exists' has been coded uniformly. For the categorical classification of the average gross wage, the category '99. Employee group not exists' has been coded.
		2012	Question pb017 ("Number of employees leaving the establishment") was not asked in 2012. However, the question is the basis for the coding of the category 'Employee group not exists'.	The coding of the category 'Employee group not exists' was carried out in 2012 without the information on employees who left the establishment. Cases containing a '-1 filter' - i.e. not coded in 'No employee group' - were coded as '-6 don't know/no answer' because according to the questionnaire there are no filters for this question.
be026_qual_imp	Average gross wage in €: employees with qualified tasks imputed	2012	The variable contains only imputed values and no original values.	In the variable all '-1 filter' values were replaced by the original values.
be026*k	Average gross wage, grouped	2013, 2014	The category '-1 filter' is not uniformly coded.	All cases that have valid metric values are set to '-1 Filter'.
		2012 to 2014	The question of the collective agreement is only put to new establishments.	For the years 2012, 2013 and 2014, the variable has been filled with the information from the previous year.
be032	Sectoral collective agreement or an in-house or establishment collective agreement	From 2015	With the 2015 wave, the question of the collective agreement allows multiple answers.	The two variables for b032_1 and b032_2 have been combined into one variable. For this purpose, the additional category '4. branch and house/establishment collective agreement' was coded.
		2011	The question of the collective agreement will be raised in 2011 by two questions.	The two variables a73 and a74 are combined in be032.

Variable(s)	Label	Year(s)	Differences	Procedure
be034	Works council or staff council elected	2012 to 2014	The question of the works/staff council is only put to new establishments.	For the years 2012, 2013 and 2014, the variable has been filled with the information from the previous year.
be035*	Chamber area	All	The items queried are not uniform over the years.	The categories '96. item not queried until 2016' were coded.
be035_1; be035_2; be035_7	Chamber area	2012 to 2014	The question of chamber membership is only put to new establishments.	For the years 2012, 2013 and 2014, the variable has been filled with the information from the previous year.
		2011	The coding of the possible answers is '0 no' instead of '2 no'.	The coding is adjusted '0 no' becomes '2 no'.
be035_7	Chamber area: other	From 2016	With the 2016 wave, the chamber affiliation was recorded more precisely, i.e. more categories were queried and the item "Other" was omitted.	If one of the variables from be035_3 to be035_6 contain the value '1 yes', be035_7 is recoded as '1 yes'. If all variables have the value '2 no' or once an invalid value, be035_7 is encoded as '2 no'.
be041	Year establishment was founded	From 2011	The question is only put to new establishments.	For panel establishments, the variable is filled with the information from the previous year. All cases that still have the value '-1 filter' are coded in '-6 don't know/no answer' because the questionnaire states that the question should not have a filter.
be043	Is your business...	2012 to 2014	The question of the type is only put to new establishments.	For the years 2012, 2013 and 2014, the variable has been filled with the information from the previous year. All cases that still have the value '-1 filter' are coded in '-6 don't know/no answer' because the questionnaire states that the question should not have a filter.
<b>BIBB Qualifizierungspanel_11bis17_long2_gwa</b>				
<b>Topic: VET</b>				
ab086	Share of productive activities during presence in the establishment	2011, 2012, 2013	In cross section, the category VET year not filled' was not coded uniformly.	The category '999. VET year not filled' has been coded uniformly.
ab087*	Perceptual performance level of apprentices	2011, 2012, 2013	In cross section, the category 'VET year not filled' was not coded uniformly.	The category '999. VET year not filled' has been coded uniformly.
<b>Topic: Staff</b>				
ps039	Employees with a bachelor's degree as the highest degree	2013	The category '9999. Employee group not exists' is not coded. The corresponding cases contain a 0.	For all cases where ps030_uni (number of employees with a university degree) has a 0, variable ps039 is coded as '9999. Employee group not exists'.

Variable(s)	Label	Year(s)	Differences	Procedure
<b>Topic: Continuing education</b>				
wb017_einf* wb017_qual* wb017_hoc*	Forms of continuing training by employment group	2012, 2014, 2015, 2017	The category 'Employee group not exists' is not coded uniformly. In 2012, the category does not exist in the cross-sectional data.	The value '99. Employee group not exists' was coded uniformly for all years. The coding of the category ' Employee group not exists' was carried out in 2012 without the information on employees who left the establishment.
<b>Topic: Special modules</b>				
mo001*	Frequency of tasks	2012, 2014, 2015	The category 'Employee group not exists' is not uniformly coded. In 2012, the category does not exist at all, the cases contain the value '-1. filter'.	The value '99. Employee group not exists' was coded uniformly for all years. The coding of the category 'Employee group not exists' was carried out in 2012 without the information on employees who left the establishment. Cases containing a '-1 filter' - i.e. not coded in ' Employee group not exists ' - were coded in '-6 don't know/no answer' because according to the questionnaire there are no filters for this question.
		2012	In 2012, the answer scale is: '1. often'; '2. sometimes' and '3. never'. In 2014 and 2015, the answer scale is: '1. never', '2. very rarely'; '3. rather rarely'; '4. rather often'; and '5. very often'.	The response scale for 2011 has been adjusted and now looks like this: '1. never'; '6. sometimes'; '7. often'.



Federal Institute for Vocational Education and Training  
Robert-Schuman-Platz 3  
53175 Bonn

Phone +49 228 107-0

Internet: [www.bibb.de](http://www.bibb.de)  
Email: [zentrale@bibb.de](mailto:zentrale@bibb.de)