RESEARCH DATA CENTRE (FDZ)
of the German Federal Employment Agency (BA)
at the Institute for Employment Research (IAB)

# FDZ-METHODENREPORT

Methodological aspects of labour market data

## **07|2021 EN** Technical Report on the IAB Establishment Panel — Wave 28 (2020)

Sebastian Bechmann, Nikolai Tschersich, Peter Ellguth, Susanne Kohaut, Christina Florian

Bundesagentur für Arbeit

KANTAR

# Technical Report on the IAB Establishment Panel

## Wave 28 (2020)

Sebastian Bechmann (Kantar)
Nikolai Tschersich (Kantar)
Peter Ellguth (IAB)
Susanne Kohaut (IAB)
Christina Florian (Kantar)

Munich, June 2021
Ag 316310176

# Content

# 1 Changes Compared to 2019

- Some questions are asked not annually but every two or more years. Within the framework of this modular system a number of questions were removed from the questionnaire in wave 28 and others were included. Outside the modular system a number of new questions have been included (cf. section 3.1 in detail). This year, these were questions about the impact of the Corona pandemic.
- Against the background of the Corona pandemic, the survey was generally conducted by telephone. However, a face-to-face interview was possible if the establishment preferred this interview mode. After the conversion of the refreshment sample in the 2019 wave, the continuer sample was converted to a computer-assisted survey mode in 2020. For a detailed description, see chapter 4.2.
- Since the end of May 2018, the EU Data Protection Regulation (GDPR) has been directly applicable in all EU member states. Against this background, in the 2019 wave, the consent of the establishments was collected at the end of the questionnaire for a renewed contact in the follow-up wave in 2020 by Kantar (Art. 6 para. 1 a, DSGVO). For this reason, the sample of panel establishments was approximately 10 percent smaller than in the 2018 survey year. This loss was offset by an increased gross input in the sample of establishments surveyed for the first time. For detailed information on the sample, see chapters 2 and 4.2. 2020 this consent question was no longer asked as agreed with the IAB.
- Since the 2020 wave, in conclusion with the IAB, the establishments have been asked at the beginning of the questionnaire for their consent to link the data collected in the questionnaire with data available at the IAB for research purposes, as the explanations in the data protection sheet on the dissemination and linking of the data collected are not sufficient as a basis for the linking.

# 2 Population, Sampling Frame, Sample

## 2.1 Population and Selected Population

The population of the IAB Establishment Panel comprises all establishments with at least one employee subject to social insurance contributions as of the reference date 30 June of the previous year.[1] The basis for sampling is the Federal Employment Agency establishment file. This contains all the establishments that in the context of the social security registration process notify the social security agencies of their employees who are subject to social insurance contributions, and are given an establishment number. As of the reference date 30 June 2019 the establishment file contained 2,140,691 establishments belonging to the population, with a total of 33,338,412 employees subject to social insurance contributions. Establishments without employees subject to social insurance contributions, for example one-person establishments or establishments only with marginal employment or employing only civil servants, are not covered by the IAB Establishment Panel. For this reason, for example, with 45,092,000 people in gainful employment (Destatis 2020), the national accounts exhibit distinctly more people employed than the IAB Establishment Panel with 41,257,641 employees.[2]

## 2.2 The Logic behind the Establishment Number

An establishment denotes a regionally and economically separate unit with employees, which is awarded its own establishment number in accordance with certain rules during the registration process for the social security agencies.[3]

- Branches of one employer in different local authority districts strictly receive their own establishment number.
- Branches of one employer within one local authority district are merged under one establishment with one establishment number provided if they exercise the same economic activity. Branches with different economic activities are given different establishment numbers.

1 Private households and exterritorial organisations have been excluded since the 2004 survey.
2 Situation as of June 2019 resp. 30th June 2019.
3 For the logic behind establishment numbers and the rules on issuing them, cf. Fritsch/Brixy 2004 and further information on the website https://www.arbeitsagentur.de/web/content/DE/Unternehmen/Sozialversicherung/Betriebsnummernvergabe/index.htm.

- An establishment itself can have several establishment numbers; this applies in particular to larger establishments with different functional areas that are considered as independent concerning their administration.
- Crossholdings between companies play no role. Every legally independent company is given establishment numbers according to the rules just mentioned.

Establishment numbers are (re-)allocated when

- the establishment previously did not have an establishment number (usually because it is the first time that the establishment has an employee who is subject to social insurance contributions),
- the establishment's economic activity has changed or
- there is a change of ownership.

## 2.3 Stratification Matrix and Partial Samples

The sample is disproportionally stratified by the size of the establishment, sector and federal state. On the one hand the target degree of completion of the individual cells is determined by the scope of the basic sample and that of the extension samples specific to the federal states and sectors. On the other hand, the individual cells are drawn according to the principle of optimal stratification proportional to the number of employees. For these reasons, large establishments, small federal states and small sectors as well as the manufacturing industry in East Germany are overrepresented in the sample. These disproportionalities are corrected with the aid of a weighting procedure afterwards (cf. in detail Chapters 7 and 8). Table 11 and Table 12 in the Appendix provide an overview of the classification of the sectors and establishment sizes. The last major changes to the stratification matrix were undertaken in 2009 in respect of the delineation of sectors, during the changeover from the economic sector classification WZ2003 to WZ2008. Since 2007 East and West Berlin have been combined.[4]

There are three partial samples in total:

- **Continuer sample:** This comprises all the establishments that are willing to participate and have a valid interview from the previous year. The continuer sample reflects the panel nature of the IAB Establishment Panel. It is necessary so that panel evaluations extending beyond pure time-series analyses can be undertaken. Panel analyses examine the developments at establishment level over time. On the other hand, cross-sectional data from at least two points in time are sufficient for time-series analyses.
- **Follow-up sample:** This comprises all the establishments that are willing to participate and have a valid interview from the year before last.[5] This partial sample increases the number of cases that can be evaluated cross-sectionally.
- **Supplementary and extension sample:** This includes extension samples that are specific to federal states, and a sector-specific extension for the manufacturing industry

4 At the wish of the Halle Institute for Economic Research (IWH), Berlin was excluded from the extension to the manufacturing industry in East Germany. Since then this extension has only included East German federal states.

5 These are generally establishments which expressed the wish to miss a year, but would then be willing to participate again in the subsequent year.

in East German federal states. It also comprises establishments with a new establishment number. The aim of this partial sample is to replace panel attrition, to achieve the sample sizes required in the individual federal states and in the manufacturing industry, and to illustrate economic structural change. The attribute “New Establishment Number” has to be used carefully and can not simply be put on the same level with „New Establishment”.[6]

Up to 2002 unit-non-responses were generally excluded from further sampling processes. In the greater size classes in particular, there were ultimately hardly any new establishments in the sampling frame which might have been included in the sample. Since then, unit-non-responses can be drawn again after a three-year rest period. Establishments of this kind that are newly drawn are reincluded in the sample with a new identification number.

6 New establishment numbers are characterised by the fact that as of the reference date of the current survey wave (= 30 June of the previous year) they had at least one employee subject to social insurance contributions, whereas a year earlier they had none. This definition is aligned with the system used for the Federal Employment Agency establishment file, and is only suitable to a limited extent for identifying the genuine founding of new establishments. The establishment might have existed before as an establishment without employees subject to social insurance contributions. It also happens that establishments do not continuously have employees; these so-called perforated establishment numbers can appear in the sample of a survey year as new establishment numbers under the above definition (cf. also the overview of how establishment numbers are awarded in Section 1.2). For this reason, this characteristic should be used with care, and only when additional characteristics from the questionnaire are taken into account, such as information on the manner in which they were founded.

# 3 Questionnaire and Pretest

## 3.1 **Questionnaire Development and Module System**

The development of the questionnaire also takes account of the panel character of the IAB Establishment Panel. In every wave so-called basic modules are used with as many unchanged questions as possible. These basic information on establishments are available for every year since 2008.[7] The basic modules are supplemented by additional modules, in which more in-depth questions are asked at defined intervals of time, usually in a two-year cycle.

Every year there are also one or more focus areas for questions which are coordinated with the various clients (cf. Appendix). Here more in-depth questions are asked about particular topics or current labour market trends.[8] Table 1 below lists the questions that were included in the questionnaire besides the module system. In the development of the questionnaire itself, a compromise must be found between comparability over time and the adaptation or modification of existing questions as well as the inclusion of new and current topics.

7 Up until the 2007 wave this statement only applies to particular topic areas. Cf. in more detail Fischer et al. (2009: 138).
8 An overview of the individual questions and the questionnaire can be found in the tools for the IAB Establishment Panel at-http://fdz.iab.de/.

## Table 1: Questions (re)included in the questionnaire in 2020[9]

| *Question* | *Section* | *Question text* | *Response options* | *Last surveyed (question)* | *Changes compared to last survey* |
|---|---|---|---|---|---|
| - | One request at the beginning | The informative value of this survey can be significantly increased if we link your information for research purposes with data available at the Institute for Employment Research. Of course, all data protection rules will be observed here as well. Do you agree with this? | yes/no | first raised | - |
| 1a | Impact of the Corona pandemic | Had or has the Corona pandemic had a negative economic impact on your establishment/ office? | yes/no | first raised | - |
| 1b | Impact of the Corona pandemic | Which of the following negative impacts did or does the Corona pandemic have on your establishment/ office? For each item, please indicate whether it applies or not. | yes/no<br>open text entry of "Other, namely ..." | first raised | - |
| 1c | Impact of the Corona pandemic | How much was or is your establishment/ office negatively affected due to the Corona pandemic? Please say it using this scale! | Scale of 5<br>(1 = slight/5 = very strong) | first raised | - |
| 2 | Impact of the Corona pandemic | Has the Corona pandemic resulted in a temporary or partial closure of your establishment/ office? This does not refer to officially ordered plant closures. | yes/no | first raised | - |
| 3a | Impact of the Corona pandemic | Is your establishment/ office threatened in its existence by the Corona pandemic? | yes/no | first raised | - |
| 3b | Impact of the Corona pandemic | Are you already on the verge of going out of business or insolvency or have you already filed for insolvency? | Multiple answers | first raised | - |
| 4 | Impact of the Corona pandemic | Due to economic difficulties caused by the Corona pandemic, have you claimed financial assistance from the federal state or local governments, such as grants, one-time payments, tax deferrals or loans? This does not refer to short-time allowance from the Federal Employment Agency. | yes/no | first raised | - |
| 5 | Impact of the Corona pandemic | Overall, how satisfied are you with the policy response to the Corona pandemic? Please say it using this scale! | Scale of 5<br>(1 = very satisfied/5 = very dissatisfied) | first raised | - |

9 Questions included or dropped due to the module system are not shown in the table, but only questions are shown that are included in individual waves (or for short periods of time).

| *Question* | *Section* | *Question text* | *Response options* | *Last surveyed (question)* | *Changes compared to last survey* |
|---|---|---|---|---|---|
| 6 | Impact of the Corona pandemic | Various personnel policy measures are mentioned below. Please indicate in each case whether these measures are being or have been implemented in your establishment/ office in response to the Corona pandemic. | yes/no | first raised | - |
| 7a | Impact of the Corona pandemic | Has there been or is there an increased demand for products or services as a result of the Corona pandemic in your establishment/ office? | yes/no | first raised | - |
| 7b | Impact of the Corona pandemic | Please estimate: How strong was or is this increased demand? Please say it using this scale! | Scale of 5 (1 = slight/5 = very strong) | first raised | - |
| 8 | Impact of the Corona pandemic | Have you or have you not changed your product or service offering or the distribution channel of your products or services considering the Corona pandemic? Or are such changes not possible? | yes/no | first raised | - |
| 20a | Staff structure | Have you had or do you have short-time work since the start of the Corona pandemic? | yes/no | first raised | - |
| 20b | Staff structure | In which months of 2020 have you used short-time work due to the Corona pandemic?<br>*If short-time work was used:* For how many employees? For all months after today's survey date, estimated data will also be sufficient. | Multiple answers/ Number (numeric) | first raised | - |
| 21 | Staff structure | Have you or had you topped up the short-time allowance for all employees or only for some of the employees on short-time or not at all? This refers to top-ups not required by law. | Single Choice | first raised | - |
| 35 | Business policy and business development | Have you implemented, initiated or planned measures to change the supply structure or value chain as a result of the Corona pandemic? | Multiple answers | first raised | - |
| 56c | Vocational training and apprenticeships | And for how many of these filled training positions is the training allowance below the new minimum training allowance of €515 per month? | Number (numeric) | first raised | - |

| *Question* | *Section* | *Question text* | *Response options* | *Last surveyed (question)* | *Changes compared to last survey* |
|---|---|---|---|---|---|
| 65b | Wages and salaries | Are you oriented to an industry collective agreement with regard to wages and salaries?<br>Does this refer to … | Single Choice | first raised | - |
| 65c | Wages and salaries | Are you paying for the employees concerned (on average) … | Single Choice | first raised | - |
| 66 | Wages and salaries | What about the average agreed weekly working hours? Does this correspond exactly to the industry collective agreement or is it shorter or longer? | Single Choice | first raised | - |
| 67 | Wages and salaries | What about other essential regulations of the industry collective agreement? The listed regulations may be exceeded, adopted exactly or with certain deductions, or may not play any role for your establishment. | Scale of 3<br>(1 = exceeded/<br>2 = exactly adopted/<br>3 = taken over with some modifications)<br>Does not matter | first raised | - |
| 72 | Company further education | Did you conduct training and development measures via e-learning (e. g. webinars) in the 1st half of 2020? | yes/no | first raised | - |
| 73 | Company further education | Were these conducted via e-learning due to the exit restrictions caused by the Corona pandemic? | yes/no | first raised | - |
| 74a | Company further education | Did you have to cancel planned training measures due to the Corona pandemic? | yes/no | first raised | - |
| 74b | Company further education | How many persons were affected by canceled training and further education measures? | Number (numeric) | first raised | - |
| 82 | Operational working hours | In the course of the Corona pandemic, has your establishment/ office introduced or expanded the option of working from home (home office)? | yes/no | first raised | - |

## 3.2 Cognitive Pretest

Since the 2005 wave, there has been a cognitive pretest for the sake of quality assurance. New questions are initially checked as to whether they appear suitable in principle for an establishment survey, and whether the establishments are likely to be capable of providing information on the content being asked, to the effect that the desired information is available at establishment level. When these criteria are met, the questions undergo a cognitive pretest, for which interviews are carried out in over 100 establishments from different sectors and establishment size classes. The pretest interview consists of two parts. Firstly the test person is asked to answer the questionnaire, then the second stage comprises the actual cognitive test. The establishments are requested to comment on each question, and are asked whether they had any difficulties in answering. The establishments' ability to provide information and their difficulties in answering as well as their problems in interpreting the questions are of particular interest. The results of the cognitive pretest flow into the further development of the questions: Which questions can be included in the next wave without making changes? Which questions still need to be adjusted and which questions are too complex or not suitable for an establishment survey? (cf. in more detail Ellguth et al. 2014: 30) The cognitive pretest interviews are undertaken by specially trained project staff from the so-called "Stützpunktagenturen" [*supporting agencies*] (ProIAB) (cf. Winters/Kargus 2012).

## 3.3 Introduction of a computer-assisted questionnaire

Since 2018, the questionnaire is also available as a computer-assisted questionnaire that reflects the functionality and flexibility of the paper questionnaire. 2019, the refreshment sample was converted to a computer-assisted survey mode, followed by the continuer sample in 2020 (cf. also the notes on study design in section 4.1). The core functionalities include the following:

- The computer-assisted questionnaire can be used both in CAPI mode and in CAWI mode.
- The interviewers are thereby able to conduct the interview on site with their CAPI laptop.
- Furthermore, the questionnaire can be left to the establishment for partial or complete completion via internet (CAWI).
- The computer-assisted questionnaire makes it possible to jump to specific questions or question modules with pinpoint accuracy. The target person is able to browse back and forth like in a paper questionnaire.
- Several thematically related questions are displayed simultaneously on one screen. Filtered questions are not hidden but only deactivated. This is to avoid mode effects compared to the paper questionnaire.
- The establishments are able to archive the completed questionnaire as a PDF or printed document for internal documentation purposes.

# 4 Study Design and Field Organisation

## 4.1 **Study Design**

Due to the Corona pandemic, the interviews should preferably be conducted completely as a computer assisted telephone interview.Only at the explicit request of the target person and the interviewer, the interview could be conducted face-to-face. The questionnaire asks for a series of items of operational information (e.g. turnover, investments, total wages and salaries), about which in principle the establishment ought to be capable of providing, but which cannot always be spontaneously recalled by the person being interviewed. In such cases the interviewer can leave the questionnaire at the establishment or as a CAWI questionaire for self-completion. The person being interviewed can look up the missing information and then complete the questionnaire. At the request of the establishment the questionnaire can also be left for total self-completion. In both cases, it is the interviewer's task to collect the completed questionnaire from the establishment or to ensure that the questionnaire is finalized in CAWI. The majority of establishments are interviewed by telephone (cf. also Table 5 in section 5.2). The option of self-completion (partly or fully) is taken up most frequently by larger establishments.

This approach leads to a response rate that is stable at a high level and ensures a high data quality (cf. chapter 5 and 6): this is because the option of self-completion often results in more precise information than in purely face-to-face interviews with a higher number of missing values in terms of "don't know" or very approximate estimates. In turn the use of trained interviewers leads to fewer errors in the completed questionnaires in principle compared to the purely mail questionnaires, and the proportion of missing information is lower.[10]

The interviews are undertaken exclusively by interviewers from the in-house interviewer staff at Kantar. The interviewers are responsible for the following sub-steps:

- Contacting the establishment (in person or by phone),
- identifying a target person who is able to provide information, and motivating this person to participate,
- checking that the correct establishment unit is being surveyed[11] and
- conducting the interview.

10 Up to and including 2014, in Saarland (since 2001) and Schleswig-Holstein (since 2002) a part of the sample was conducted as a mail survey. This had become necessary for capacity reasons, because due to the extension samples specific to these individual federal states, the desired sample sizes could not have been achieved in the stipulated field period. In 2015 Schleswig-Holstein switched over completely to a face-to-face data collection mode, and Saarland followed in the 2016 wave.

11 Cf. in detail Section 4.2.

Continuity of the interviewers used is crucial for the success of the survey. In this respect, Janik and Kohaut (2009) were able to demonstrate the very strong effect a change of interviewer has on the likelihood of participation. The probability of non-participation increases markedly when a different interviewer is used to the one who was used the previous year. Changes are necessary in principle when interviewers have left in the meantime or establishments have moved. Table 2 shows that changes of interviewers are rather rare in the context of the IAB Establishment Panel and that the bulk of interviews are undertaken by the same interviewers as the previous year.

**Table 2: Summary of the use of interviewers**

| | 2020 | 2019 |
|---|---|---|
| Proportion of interviewers used again | 93 % | 82 % |
| Proportion of cases without a change of interviewer | 86 % | 87 % |
| Number of interviewers used | 419 | 496 |

The data collection generally takes place annually from the end of June to the end of October. In parallel the data that has been gathered undergo checks and errors are eliminated (for this cf. Chapter 6). In advance the establishments receive an announcement letter from the Federal Employment Agency (BA), a letter of recommendation from the Confederation of German Employers' Associations (BDA), a data privacy declaration and a summary of the most important results from the previous year – in the form of a 8-page flyer.

## 4.2 Field organization

In the 2018 wave, as part of a mixed-mode experiment, the sample of establishments to be surveyed for the first time was divided into different subsamples and successively surveyed using different modes (see chapter 5.3 of the 2018 technical report). The establishments in this experiment were surveyed in 2019 using the same survey mode as in 2018.

The refreshment sample was almost entirely switched to a computer-assisted questionnaire in 2019. Further information can be found in the method report 2019.

In view of the Corona pandemic, the IAB and Kantar have decided to redesign the field organization in such way that the IAB Establishment Panel could be conducted even under very restrictive conditions. For example, at that time it was not foreseeable whether a contact restriction due to legal requirements would prevent the IAB Establishment Panel from being conducted face-to-face during the period scheduled for the field work. In addition, it was unclear whether a face-to-face interview situation was at all justifiable for the interviewers and the target person against the background of the infection process. Therefore, the following basic changes in the field organization were agreed upon for the 2020 wave:

- Interviewer should perform the contacting of the establishment and, as far as possible, the interview by telephone. The interviewers should possibly transfer the information provided by telephone directly into the computer assisted instrument.

- If the interview was conducted by telephone, the interviewers should ensure that the target person had access to the questionnaire. This could be done by sending the paper questionnaire with the invitiation letter, by accssing the website iab-betriebspanel.kantar.com or by a questionnaire sent ad hoc by the interviewer (by post or by e-mail as a PDF).
- Alternatively, the interviewers could also motivate the target person to complete the questionnaire themselves, either online in the CAWI instrument or written via paper questionnaire.
- Interviews were also allowed to be conducted face-to-face in case of an explicit request of the target person (and with the consent of the interviewer). This option was especially kept for establishments of the continuer sample, which would otherwise possibly refuse to participate in the IAB Establishment Panel.

Accompanynig measures:

- Establishments from the continuer sample received an announcement letter some time before the invitation letter, which was intented to inform about the changed procedure of the survey.
- The refreshment sample was divided in three groups: the first part was given to the permanent interviewers of the Kantar staff for telephone processing right at the beginning, the second part was initially set up as a self-completion sample and only handed over to the interviewer staff for processing after a reminder, and the third part was a pure self-completion sample. The background to this procedure was that the initial respondent sample in this wave was very large. The fact that part of the sample was processed by the interviewers right from the start meant that the addresses could be processed more efficiently, as the field time would otherwise have been too short.
- Large parts of the sample were created as a CAWI-1$^{st}$-sample. The establishments received a link to the CAWI questionnaire with their invitation letter as well as a paper questionnaire with the option to complete the questionnaire themselves. After a certain period of time, the companies were contacted by the interviewers if no interview was available.

The refreshment sample consisted of three subgroups, the continuer sample of six subgroups. In overview 3 a detailled summary over the single subgroups can be found, with information on

- the survey mode
- the contents of the individual letters
- the mailling dates
- the start of the processing by the interviewers

## 4.3 Identification of the Correct Establishment Unit

The interviewer is responsible for ensuring that the correct establishment unit is surveyed. For establishments being surveyed for the first time, the establishment number defines the survey unit. The interviewer has to decide on the spot whether the information requested in the questionnaire is actually available for this unit. The establishment number itself is a characteristic that is difficult to handle when identifying the correct establishment unit. For that reason the interviewers – aside from the name of the establishment – rely on the sector and the number of employees subject to social insurance contributions as of the reference date of the previous year, according to the details held by the Federal Employment Agency (BA). It is only in the event of deviation that the interviewers resort to the establishment number. In specific cases the interviewer can interview a different unit, especially if the unit specified by the establishment number does not constitute a unit which is worthwhile interviewing.[12] If there is another possible survey unit, there must still be some relationship to the original establishment number on the address protocol: The deviating survey unit must contain all or part of the given establishment number. If this is not the case, no interview takes place.

In the case of establishments periodically interviewed, the same unit as the unit in the previous year should be interviewed. This applies irrespective of the relationship between the unit surveyed the previous year and the originally stipulated establishment number. The connection to the previous year is established using the details of the total number of employees and sector from the previous year. Only when the provided information in the current year on the reference date of the previous year are identical with the corresponding information given in the previous year, it can be assumed that the interview is taking place in the same establishment as last year (in terms of the employee numbers within a defined range of tolerance).[13] This information forms the basis of the panel case definition provided, and thus of the individual longitudinal sections (cf. Chapter 8 on this in detail). It should be pointed out that as well as this panel case variable, the researcher should always also make use of additional variables to define establishments that are identical with the previous year for his specific interest. Thus a decision on whether, for example, an establishment which has closed some subunits, but has also integrated other units, is still the same establishment or not has to be taken on the basis of considerations of content.

12 This can occur, for example, if individual functional areas (e. g. administrative and care personnel in a hospital) are each given their own establishment number, or the establishment number comprises several branch offices (cf. Section 2.2 in this respect too). This is found in particular in the public sector, the non-profit sector, manufacturing industry with a relatively large proportion of large establishments, and in sectors with a large number of legally independent branch offices such as retail. According to Fritsch/Brixy (2004: 185f), this occurs in less than ten percent of cases, and discrepancies in terms of numbers of employees are generally relatively small.

13 For establishments with more than 20 employees the range of tolerance is +/– 5 %, for establishments with up to 20 employees +/– 1.

## 4.4 Training and Monitoring Interviewers

Kantar continuously recruits interviewers for its in-house f2f interviewer staff. Interviewers are selected in an extensive, multistage selection and assessment process, within the prospective interviewers receive training face-to-face and in writing. They receive detailed basic information about statistical selection procedures, data protection and interviewing behaviour. The so called contact interviewers, experienced interviewers who take over the local supervision of the new interviewers, practise the later interview situation with the new interviewers. Since contact interviewers also work as interviewers, they are able to pass on their experiences and practical tips to the new interviewers. The new interviewers are intensively supervised by the contact interviewers during their first projects. The interviewers at Kantar receive follow-up training as standard at regular intervals.

As already mentioned above, in 2020 almost 93% of the interviewers deployed for the IAB Establishment Panel had already worked for the project in one or more previous waves. Thus they are extremely familiar with the special requirements of the project. Extensive written instructions for interviewers were also prepared for the current survey wave, dealing with the following points: new features compared with the previous wave, the issues being studied, clarification of the correct establishment unit, information about making contact and documenting contacts, identifying the main target person(s) for the editing, information about data protection and special information about the questionnaire as far as necessary.

In addition to these written instructions, interviewers working for the IAB Establishment Panel for the first time also receive personal project training from the responsible contact interviewer. Kantar has prepared a project-specific concept and corresponding training documents especially for this. During this training, the contact interviewers convey the most important information about the project, such as the client, duration and scope of the project, the background and objectives of the study, and what the collected data will be used for. They explain to the interviewers how they should identify the correct establishment unit and the target person, and they are told the importance of ensuring that the establishments take part. Finally the new interviewer and the contact interviewer go through the questionnaire together with the contact interviewer explaining it.

As well as the intensive interviewer training for quality assurance purposes the interviewers' work is continuously monitored to ensure "that no significant falsification of the research results takes place through interviews – consciously or unconsciously – not being conducted correctly" (Sommer et al. 1999: 414):

- For the IAB Establishment Panel a project-specific control of all the interviews that have been conducted takes place through the extensive ex-post data validation and follow-up telephone surveys (cf. in detail Chapter 6). In order to make sure that interviews that are correct right from the start will be considered for checks, an additional random-based control was inroduced. The interviews of the interviewers with more than 50 interviews are checked with a control rate of at least 25 % each. The interviews of the interviewers with up to 50 interviews are checked with a control rate of at least 10 % each.
- Beside this project specific control in the IAB Establishment Panel, the interviewers' work is also randomly checked in other projects using random sampling procedure. The interviews of those interviewers that were conspicous in the context of these control measures were included in the project-specific control of the IAB Establishment Panel (see first bullet point).

- Thus amongst other things, the interviewee stated in each case is subsequently asked by mail, by phone or also face-to-face whether, when and with whom the interview took place and how long it lasted. If the checks give cause to doubt the quality of an interviewer's work, this interviewer is asked to provide a written comment. The interviewer is given further training as appropriate or – in serious cases – he runs the risk of reduced fees, or even of leaving the interviewer staff. If an interviewer becomes conspicious during a routine check, also his faultless questionnaires get checked in project specific controls as mentioned above.
- In addition, interviews were placed in the control just described if they were conspicuous in the so-called similarity check. This check is based on the hypothesis, which is also supported by experience, that interviewers who 'fill out questionnaires themselves' generate data with less variance than in real interviews. Statistical methods are used to test whether there are greater similarities between the interviews of one and the same interviewer than between the interviews of other interviewers. A low variance is only a first indication that the interviewer has falsified data. Therefore, such interviews are given to the telephone follow-up check described above. In order to avoid biased results during the similarity check, variables are removed from the data set before the analysis which a) are within strong filtering by the question program, b) are by definition more similar within an analysis stratum than between analysis strata and c) contain a relatively high proportion of missing values. In addition, the similarity check is carried out separately for individual establishment size classes, since smaller companies by definition have a lower variance in their data than larger companies and otherwise false positive results are frequent.

  A total of 99% of the interviewers used for the IAB Establishment Panel were checked in 2020 as part of this control. For two interviewers anomalies were detected. In one case, a lower response rate than average was documented for the control letter. The interviewer was retrained that he or she should prepare the target persons for a possible postal quality control by Kantar and better document the addresses. In the other case, there was an accumulation of short interviews and poor address documentation. The interviewer was admonished and intensively retrained to better document addresses and to read the questions more slowly and verbatim. Interviews of interviewers who were conspicuous during the control were additionally checked during the editing process. However, none of the interviews conducted by these interviewers revealed any further anomalies.

# 5 Result of Field Work

## 5.1 Overview of Field Work

The field work started on 01 July 2020, with the last interview being held on 16 November 2020. Figure 1 shows both the weekly response rate in terms of the interviews held and the reported non-responses, as well as the cumulative response rate.

**Figure 1: Number of interviews and non-responses as well as cumulative response rate per week (in number of responses)**

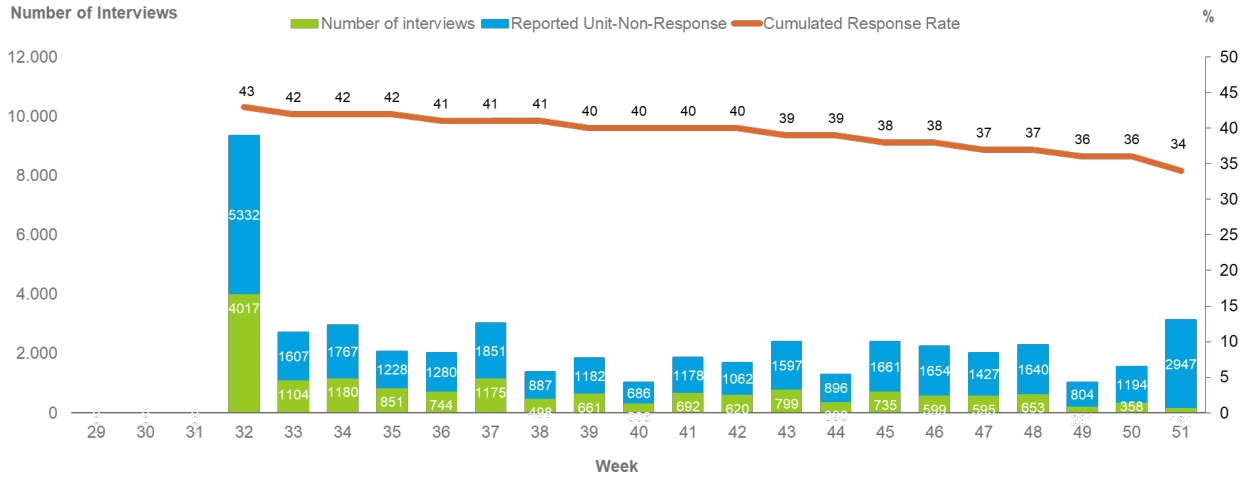



Figure 2 shows an overview of the distribution of the interviews by the date on which they were held – and not by the date of the return as in Figure 1. The typical field work progression can be seen clearly: the number of interviews achieved per week rises initially and falls continuously as the field period increases.

**Figure 2: Proportion of interviews (date on which they were held), in percent**

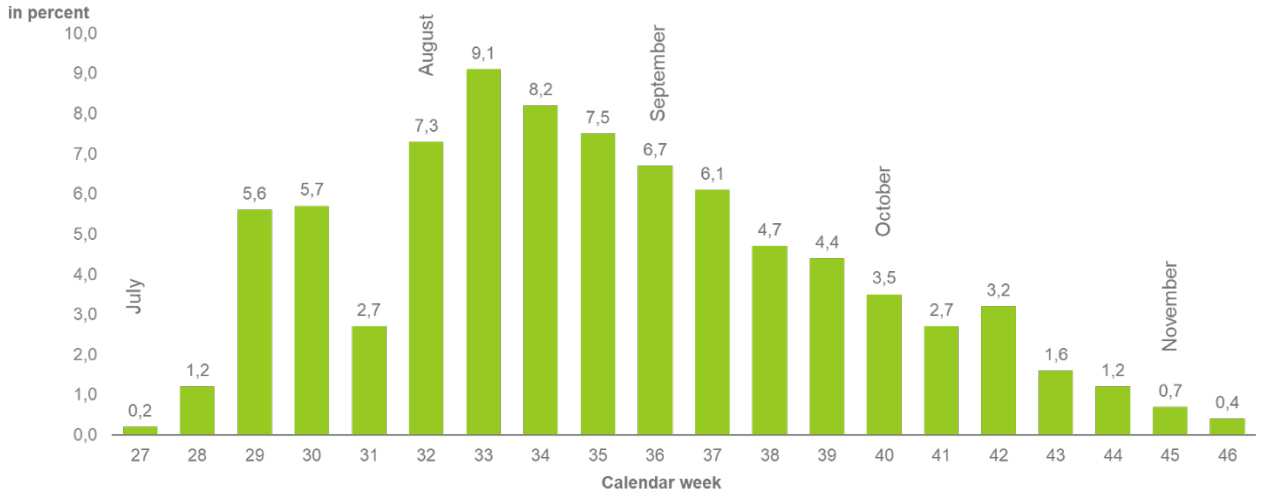
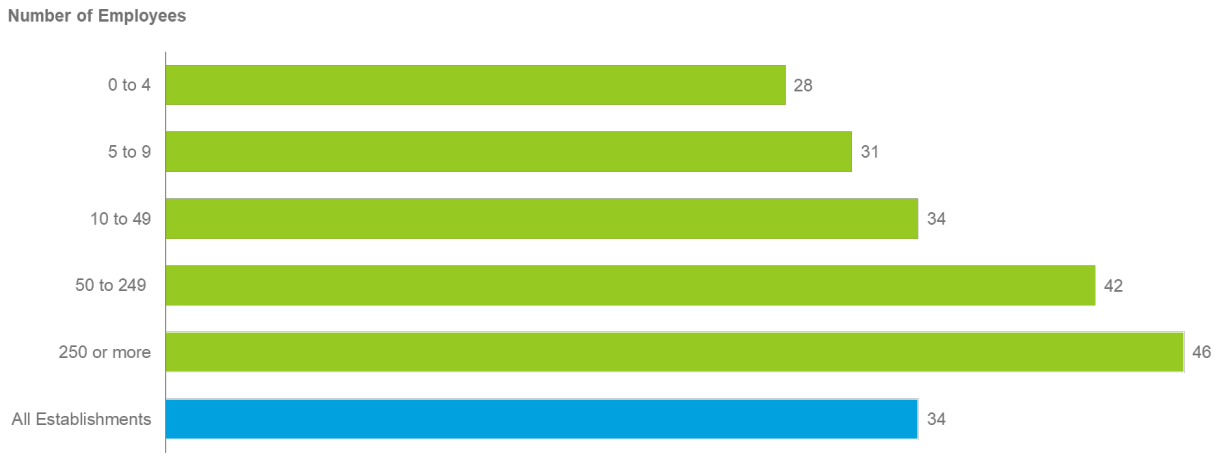



Figure 3 shows the average duration of the interviews by establishment size class. The average interview duration in smaller establishments is in principle lower than in larger establishments, mainly due to the fact that smaller establishments have a better ad hoc overview of what happens in the establishment than larger establishments. Thus questions about the personnel structure (e. g. proportion of part-time employees) are easier to answer in a 3-person-establishment than in an establishment with 100 employees. In addition, certain issues frequently do not apply to smaller establishments, so they have to answer a greatly reduced catalogue of questions due to the use of suitable filter questions. Thus, for example, the entire block of training or development questions is omitted if the establishment is not an authorised training provider or no employee received training respectively. Both of these cases occur more frequently with smaller establishments than with larger ones. As a result we can see that as the number of employees increases, the average interview duration rises from 28 minutes approximately 46 minutes. The average interview duration in 2020 was approximately 34 minutes.

**Figure 3: Average interview duration by establishment size class, in minutes**



## 5.2 The Responses in detail – Unit-Non-Response

As already described in Section 2.3, the IAB Establishment Panel is disproportionately stratified according to different characteristics. Table 3 provides an overview of the extent of the gross sample used and the numbers of cases actually realised for each federal state (federal state-specific extension samples) and for the manufacturing industry in East Germany (sector-specific extension sample). The number of establishments successfully re-interviewed is shown separately.

**Table 3: Overview of gross and net sample, by federal state and manufacturing industry in East Germany[14]**

| | Gross | Net (actual) | of which repeated |
|---|---|---|---|
| Schleswig-Holstein | 2,027 | 832 | 648 |
| Hamburg | 1,301 | 291 | 158 |
| Lower Saxony | 2,929 | 1,133 | 831 |
| Bremen | 2,009 | 869 | 663 |
| North Rhine-Westphalia | 7,644 | 1,904 | 1,006 |
| Hesse | 3,722 | 1,046 | 663 |
| Rhineland-Palatinate | 2,537 | 864 | 615 |
| Baden-Württemberg | 3,936 | 1,258 | 802 |
| Bavaria | 5,632 | 1,478 | 773 |
| Saarland | 4,687 | 1,186 | 579 |
| Berlin | 4,598 | 1,054 | 542 |
| Brandenburg | 2,326 | 1,049 | 843 |
| Mecklenburg-West Pomerania | 2,147 | 936 | 741 |
| Saxony | 2,547 | 1,137 | 902 |
| Saxony-Anhalt | 3,280 | 1,115 | 736 |
| Thuringia | 4,022 | 1,226 | 757 |
| **Total** | **55,344** | **17,378** | **11,259** |
| Manuf. ind. East Germany (excl. Berlin) | 3,182 | 1,559 | 1,214 |

From the total gross sample of 55,344 establishments, a total of 17,378 valid interviews were realised. Thus compared with the total sample, a response rate of about 31 % was achieved (cf. Table 4).

14 The manufacturing industry in East Germany is shown separately as these cases are already contained in the case numbers differentiated by East German federal states (excluding Berlin).

**Table 4: Number of evaluable interviews and response rate, by partial sample[15]**

| Partial samples | Region | Gross (absolute) | Evaluable interviews | |
|---|---|---|---|---|
| | | | absolute | as % of gross |
| a) Respondents from wave 2019 | West | 9,824 | 6,738 | 69 % |
| | East | 6,392 | 4,521 | 71 % |
| | **Total** | **16,216** | **11,259** | 69 % |
| | | | | |
| b) Samples of non-respondents from previous year willing to be surveyed again in 2020 | West | 726 | 136 | 19 % |
| | East | 411 | 91 | 22 % |
| | **Total** | **1,137** | **227** | 20 % |
| | | | | |
| c) Refresher sample 2020 | West | 25,874 | 3,987 | 15 % |
| | East | 12,117 | 1,905 | 16 % |
| | **Total** | **37,991** | **5,892** | 16 % |
| | | | | |
| d) Gesamt | West | 36,424 | 10,861 | 30 % |
| | Ost | 18,920 | 6,517 | 34 % |
| | **Gesamt** | **55,344** | **17,378** | **31 %** |

The different subsamples differ greatly in terms of the response rates.[16]

- In the subsample of continuer establishments (excluding non-respondents from the previous year willing to be surveyed again), at 69 % the response rate is distinctly higher than in the establishments being surveyed for the first time, at about 16 %. In the sub-sample of establishments surveyed repeatedly, which is particularly important for the project, a high response rate was once again achieved, but it is lower than in the years before 2019. As in 2019, the lower utilization is attributable to the significant increase in the number of refusal in advance, thus establishments that gave an interview in the previous wave but had already indicated before the start of the current survey that they no longer wished to participate. The introduction of an explicit question on willingness to be re-contacted by Kantar in 2018 questionnaire resulted in a methodological change. 1,531 (77.2%) of the 1,984 prior refusals were based on this question; only 453 were made by other means (by telephone or e-mail on the part of the establishments or by feedback from the interviewer) (in 2019 there were 464). In addition, the conditions made more difficullt by the Corona pandemic will have played a role.[17]
- The slightly higher response rate in the sub-sample of establishments surveyed for the first time compared to the 2019 wave is mainly due to an increase in the response rate in the CAWIonly subgroup R3 by around 3 percentage points. However, the CAPI processing was also somewhat more successful than in the previous year.

15 Due to the method test mentioned above, the exhaustion rates for the total and supplementary sample are no longer comparable with previous years.
16 Excluding the neutral non-responses (e. g. addresses for which it was not possible to determine whether they belong to the target group), a cooperation rate of 55 % was achieved.
17 The explicit question on willingness to be re-surveyed was introduced in 2018 in response to the EU Data Protection Regulation (GDPR), which came into force at the end of May 2019.

- The response rate of about 20 % in the subsample of non-respondents from the previous year willing to be surveyed again is slightly higher than that in the partial sample of establishments being surveyed for the first time.

Table 5 shows an overview of the type of implementation. The type of interview was recorded in the address protocol for each establishment. Since the 2020 wave also included interviews conducted entirely or partially by telephone for the first time, a comparison with the previous year's figures is only of limited expressiveness. The proportion of fully interviewer-assisted surveys declined again and now accounts for just under the majority of all interviews (54 % completed entirely by respondents themselves, 45 %), compared with 77 % in 2019. As in 2019, the low proportion is also attributable to subgroup R3, in which the companies completed the questionnaire exclusively themselves. If this subgroup is excluded for the calculation, the proportion of those farms that completed the questionnaire entirely themselves, at 33 %, is still significantly higher than the figure from the previous year (13 %). This may be due to the fact that the telephone survey was the preferred mode in this wave and not, as is usually the case, a face-to-face interview.

**Table 5: Overview of the form of the interview**

| | Total | | Respondents from previous wave | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | Total without R3 | P1 CAWI1st | P2 CAWI1st | P3 CAWI1st | P4 CAWI1st | P5 CAWI1st | Total |
| | | | | | | | | |
| conducted entirely by telephone | 45% | 54% | 20% | 53% | 34% | 42% | 12% | 50% |
| conducted entirely face-to-face | 9% | 12% | 4% | 15% | 20% | 13% | 0% | 14% |
| conducted partly by telephone | 1% | 1% | 1% | 1% | 2% | 2% | 0% | 1% |
| conducted partly face-to-face | 0% | 0% | 0% | 0% | 0% | 1% | 2% | 0% |
| entirely completed by the respondent himself | 45% | 33% | 75% | 32% | 44% | 43% | 86% | 34% |

| | Refresher sample (incl. temp. failures) | | | | | |
|---|---|---|---|---|---|---|
| | P0 CAWI1st | R1 CAPI | R2 CAWI1st | R3 CAWIonly | Total | Total without R3 |
| | | | | | | |
| conducted entirely by telephone | 45% | 83% | 38% | 3% | 35% | 68% |
| conducted entirely face-to-face | 9% | 1% | 0% | 0% | 1% | 1% |
| conducted partly by telephone | 2% | 0% | 0% | 0% | 0% | 0% |
| conducted partly face-to-face | 1% | 0% | 0% | 0% | 0% | 0% |
| entirely completed by the respondent himself | 43% | 16% | 61% | 97% | 64% | 30% |

As previously, the form the interview takes is substantially associated with the size of the establishment: the proportion of establishments surveyed entirely face-to-face falls linearly from 67 % of the smallest establishments (with 1 to 4 employees) to 25 % of the large establishments with 5,000 or more employees. This is because the larger the establishments the more frequently the interviewee is able to provide the complex quantitative details only with extensive preliminary work; this applies in particular to the employee structure, personnel recruitment, appointments and resignations, training, business volume and investments. In these cases it is helpful for the interviewee and the interviewer if the questionnaire can be left at the establishment for further processing or is made available to the respondent in advance (cf. also Section 4.1).

Table 6 shows the correlation already mentioned in Section 4.1, namely that a change of interviewer has a negative effect on the establishments' willingness to participate. In the group without a change of interviewer, the response rate is 4 percentage points higher than that of the group with a change of interviewer. It is noticeable that, compared to previous years, the gap between the successes in exhaustion is smaller and the exhaustion in the companies with a change of interviewer is higher than in previous years. It seems that the personal connection is less important in a survey that is mostly conducted by telephone.

**Table 6: Response rate with and without a change of interviewer**

| | Response rate |
|---|---|
| The same interviewer as previous year | 78 % |
| Different interviewer to previous year | 74 % |

## 5.3 Dealing with Missing Information – Item-Non-Response

A further element in ensuring the highest possible data quality is the way missing information (so-called item non-response) is dealt with. In the IAB Establishment Panel this includes on the one hand measures to prevent missing details, and on the other hand the most differentiated recording possible of the different forms of missing values.

Missing details especially arise because the corresponding question is difficult to understand, difficult to answer for factual reasons or the interviewee refuses to answer (e. g. for reasons of confidentiality or the time required to look up the answer). The already mentioned cognitive pretest serves not least to identify such questions that will probably be problematic in advance, and as applicable to revise them or even not take account of them at all (see Section 3.2). To some extent to avoid missing values the interviewees are given the option of stating estimated values ("If it is not possible to answer precisely, please estimate").

A particulary important factor in minimising missing information is the use of interviewers. Thus, as in previous years, the frequency of missing information has shown clear differences between face-to-face interviews and interviews conducted via mail. In general, the quality of face-to-face interviews (and also of those cases in which the questionnaire was left at the establishment for (partial) self-completion) was significantly better than interviews via mail.[18]

Questions that do not apply to an establishment (e.g. follow-up questions on further vocational training in establishments that have given negative answers to the initial filter question) are simply not asked at all in the IAB Establishment Panel. If a question does not apply to the establishment, the corresponding variable in the data set is empty (system-missing), but can of course be recoded into a corresponding missing category at any time. If however the use of filters means that for example certain groups of employees do not exist in the establishment (e.g. number of employees who have received further training), then the missing information can/must be replaced by the user with a zero. In this respect the reason for a non-response "Does not apply" is not coded separately in the IAB Establishment Panel.

Generally in literature, when designing questions for item non-response a differentiation is required between the categories "No answer/declined to answer" and "Don't know". In our

18 Thus, the average proportion of missing values for interviews conducted entirely face-to-face was at one percent, for those completed by mail at five percent.

view these categories cannot be – with few exceptions – reliably differentiated. Many items of information that are required – even if they are available in the establishment – are associated with a certain research effort, or passing them on is regarded as sensitive. In the event of a refusal to answer, the interviewees would presumably choose one of the two options depending on the topic.

Against this background, apart from a small number of exceptions the explicit provision of answer categories for “Don’t know” or “No answer/declined to answer” has been avoided in the questionnaire. If such answers are explicitly provided, this attracts such responses in addition, as it is easier to tick “Don’t know” than to search for a particular business figure.[19] As an example: up to the year 2000 the question in the IAB Establishment Panel about advance payments offered “Don‘t know” as a permissible option to response, but with the 2001 wave, while the question was otherwise unchanged, this option no longer existed. In the 2001 wave, the proportion of missing answers fell by 21 percentage points.

In a small number of cases the “Don’t know” category contains utilisable information and can be used for analysis purposes. Thus, for example, the question on the development of the volume of business expected in the current financial year (in comparison to the previous year) includes the category “Don’t know yet”. Another example concerns special labour market programmes from the Federal Employment Agency, which under certain circumstances are not yet known in the establishment. In such cases the corresponding variable in the data set is given its own code for “Don’t know (yet)/Cannot say/Not known”. There are seven such questions in the entire 2020 questionnaire (questions 42, 43, 44, 66, 77c, 88a, 88b). For all other questions there is no separate answer category “Don’t know” (or a category with similar content), which is the reason why this category cannot be displayed separately.

Overall in the 2020 survey, 24 % of all questions/variables had less than 0.5 % missing values, 61 % had less than two percent and 87 % had less than five percent missing values. Table 7 provides an overview of the questions with a very high proportion (10 % and more) of missing values. When evaluating variables with a lot of missing information, possible distortions as a result of this should of course always be borne in mind.

**Table 7: Questions with a high proportion of missing values**

| Question/ Variable | Content | Unit | 2019 |
|---|---|---|---|
| bb11 | Number of employees leaving due to closure/spin-off/spin-off | (number) | 32 % |
| bb40 | Share of intermediate inputs/external costs in sales 2019 | (%) | 29 % |
| bb37 | Business volume 2019 | (EUR) | 25 % |
| bb69 | Gross payroll June 2019 | (EUR) | 23 % |
| bb01bf | Consequence of the Corona pandemic: Other, namely | - | 18 % |
| bb18_1eu | Number of 1-euro jobbers 06/20 | (number) | 14 % |
| bb44b | Projected number of employees for the coming year in-total | (number) | 11 % |
| bb47 | Share of expansion investments | (%) | 10 % |

19 This phenomenon comes under the problem of satisficing. Detailed explanations can be found e.g. in Krosnick et al. (1996).

# 6 Data Verification and Follow-Up Telephone Interview

In parallel with the field work, the data that has already been collected is checked both cross-sectionally and longitudinally for its completeness, consistency and plausibility. In the 2020 wave, in total 161 cross-sectional checks, 25 longitudinal checks and 48 filter checks were carried out. A distinction is made between four essentially different checks:

- **Filter errors**: Were the filter instructions obeyed in the intended way, or have establishments mistakenly answered resp. not answered a question?
- **Checks for completeness**: For selected questions checks are undertaken as to whether the question was answered. This applies in particular to questions that are relevant for weighting, such as the question on the number of employees subject to social insurance contributions.
- **Plausibility checks**: These checks assess information that are generally unlikely, but can nonetheless occur in practice. One example of this type of check involves reviewing the per capita incomes. The check indicates an error for values which are comparatively high or low.
- **Consistency checks**: These checks refer to logical connections between different answers. Consistency criteria are infringed, for example, if the questionnaire contains contradictory information. The consistency checks include for instance checking the total amount stated in one question against the total of the individual values.

If missing or incorrect information cannot be supplemented or corrected by means of the questionnaire, an attempt is made to obtain clarity together with the interviewee during a follow-up telephone interview. As a result, missing information can be supplemented and incorrect information corrected. In plausibility checks, implausible values are released with corresponding justification after consultation with the establishment – so despite infringing the test conditions, the value will be accepted.

In the 2020 wave, 23 % of all interviews were completely error-free right from the start, and therefore did not require any further processing. Corrections of the remaining 77 % not completely error-free interviews were undertaken during the editing process, and in the majority of these cases (5,166 respectively 38 %) the establishment was also contacted for a follow-up telephone interview.[20] This intensive downstream data check also involves a comprehensive project-specific review of the work of the interviewers (cf. also section 4.3). A further 2,257 cases were controlled by the random-based control described in section 4.4.

20 This corresponds to 30 % of all 17,378 evaluable interviews.

# 7 Definition of Cross-Sectional Cases and Projection

## 7.1 Definition of Cross-Sectional Cases

All cases for which a valid questionnaire has been completed and which had at least one employee subject to social insurance contributions as of the reference date of 30 June of the previous year are referred to as cross-sectional cases. Due to the disproportionate structure of the sample the data has to be weighted before descriptive evaluations. With a disproportionate sample structure analyses of unweighted data lead to non-representative results.

For 2020 in total 16,686 cases are available for cross-sectional analyses, of which 10,481 are from West German federal states and 6,205 from East German federal states. 692 further interviews were conducted with establishments which had no employees subject to social insurance contributions as of the reference date of 30 June of the previous year, and therefore did not belong to the population for cross-sectional evaluations; these cases are only used for longitudinal evaluations (see also table 9).

## 7.2 Cross-Sectional Weighting Process

Weighting is necessary because the sample structure is disproportionate in terms of the establishment size, sector and federal state, and is also needed in order to compensate for any possible differences between the actual and target size of the individual stratification cells (cf. also Section 2.3).

The weighting of the IAB Establishment Panel essentially takes place in the form of a projection onto the population. This applies for the establishments in the population, in other words for the whole of Germany but also for East and West Germany, for the individual federal states, and for manufacturing industry establishments in East Germany. The target structures are taken from the establishment file of the Federal Employment Agency. The determining factor is the structure (distribution of the establishments) at the time of drawing the sample for the respective wave (in other words as per 30 June of the respective previous year). This projection compensates the disproportionalities as well as the different response rates along the stratification cells in one step.

The weighted sample of the IAB Establishment Panel is proportional to the number of establishments, and thus reflects the distribution of the establishments across the cells of the stratification matrix. Its structure therefore differs from numerous other establishment surveys, in which the over-representation of large establishments is not corrected (and which thus deliver results that are proportional to the number of employees or the turnover, but not to the number of establishments). However, the weighted data of the IAB

Establishment Panel enables analyses that are not only proportional to the establishments, but also proportional to the employees. Moreover, during the cross-sectional weighting, attention is paid to ensuring that at federal state level the employee figures projected from the weighted sample (employees subject to social insurance contributions as of the reference date of 30 June the previous year) correspond to the targets of the Federal Employment Agency. For analyses that are proportional to the number of employees, the weighted number of employees from the establishments to which the characteristic in question applies has to be set in proportion to the total number of employees.

The result of the weighting is an integrated weighting factor that is proportional to the numbers of establishments and employees:

- A weighting that is proportional to the number of establishments reflects the distribution of the establishments across the cells of the stratification matrix. This enables representative statements to be made on the percentage of the establishments in Germany to which a particular statement (e.g. establishment has a works council) is applicable.
- The weighting that is proportional to the number of employees reflects the distribution of the employees across the federal states in Germany. This enables representative statements to be made on the percentage of the employees that work in establishments to which a particular statement is applicable (e. g. employees work in an establishment with a works council).

For drawing the sample and for projections proportional to the number of establishments the population is subdivided into 19 sectors and 10 establishment size classes per federal state (see Table 11 or Table 12 in the annex). This creates a stratification matrix consisting of 190 cells per federal state. This subdivision has been in place since 2010.

For the IAB Establishment Panel the weighting factors of the cross-section are calculated using generalized regression models, GREG for short.

The aim of a random sampling, i.e. design-based inference, is to estimate certain parameters of the population for a target characteristic y of interest.[21] Important parameters are the sum or the average of this target characteristic in the population. If, in order to estimate such parameters from the population U (= 1, ..., k, ..., N), a sample s (= 1, ..., k, ... , n) with strictly positive selection probabilities for each element ($\pi k = Pr(k \in s) > 0$, $\pi kl = Pr(k \& l \in s) > 0$) based on the sample design,[22] then the design weight dk of a sample element k is the inverse of its selection probability πk, i.e., $dk = \pi k^{-1}$. The Horvitz-Thompson estimator then represents the design-weighted estimate of the parameter, so, for example, for the sum $\hat{y}$ of a feature y, the value $\hat{y} = \sum s\ dk\ yk$

In order to take into account not only design-related differences in selection probabilities but also the failure event and to reduce the variance of the estimators, the IAB Establishment Panel uses generalized regression models (GREG) to further adjust the weighting factors by adding certain auxiliary characteristics. Starting from a Horvitz-Thompson estimator, the aim of a calibration by GREG is to adjust the design weights against the background of additionally available information on the sum (or average) of auxiliary characteristics x and to convert them into new weighting factors wk in such a way that the sample after weights represents the sum (or average) of these auxiliary characteristics x, i.e. $\sum s\ wk\ xk = \sum U\ xk$.

[21] The following statements are based on Deville, J.-C., Särndal, C.-E., Sautory, O., 1993: Generalized Raking Procedures in Survey Sampling. Journal of the American Statistical Association, Vol. 88, No. 423, pp. 1013–1020.

[22] The second requirements of the strictly positive selection probability is necessary to be able to determine the variance of the estimators (cf. e.g. Cassel, C.-M., Särndal, C.E., Wretman J.H., 1977: Foundations of Inference in Survey Sampling. New York: John Wiley & Sons.)

At the same time, the original design weights dk should be changed as little as possible: "Our objective is to derive new weights that modify as little as possible the original sampling weights (dk = πk-1), which have the desirable property of yielding unbiased estimates" (Deville and Särndal 1992).[23]

The weights wk are the solution of a minimization problem under constraints: If G(wk/dk) denotes a function that maps the distance between dk and wk, then the optimization problem is to minimize the function shown below, with respect to wk, where λ represents the vector of Langrange multipliers.

**Optimization problem in the context of a generalized regression[24]**

$$\sum_{S} d_k \, G\left(\frac{w_k}{d_k}\right) - \boldsymbol{\lambda}'\left(\sum_{S} w_k \boldsymbol{x_k} - \sum_{U} \boldsymbol{x_k}\right)$$

with:

$w_k$ = final weighting factor

$x_k$ = auxiliary characteristics of the elements of the

$d_k$ = design weight

$S$ = sample

U = population

G = distance function

$\lambda$ = langrange multiplier

Deville et al. (1993) describe several distance functions. The variant described by Deville et al. (1993: 1014) as linear method leads in the application to the establishment panel sample to the best adjustments with slightly larger factor ranges in comparison to the procedure described by them as logit method. Because of the better adjustments, the linear method was therefore chosen. If one uses this method, then for the estimation of the sum of a characteristic the "generalized regression estimator" (GREG, see Deville, Särndal 1993: 1014) results:

23 Deville, J.-C., Särndal, C.-E., 1992: Calibration Estimators in Survey Sampling, Journal of the American Statistical Association, Vol. 87, No. 418, pp.376–382.

24 Here and in the following equations letters in bold denote vectors, in normal letters scalars.

**Estimator of the sum of the feature y based on generalized regression (GREG)**

$$\hat{y}_{reg} = \sum_{S} w_k \, y_k = \hat{y}_\pi + (\hat{\boldsymbol{x}} - \hat{\boldsymbol{x}}_{\boldsymbol{\pi}})' \hat{\boldsymbol{B}}_{\boldsymbol{s}}$$

Here, the estimators y and x denoted by the subscript π denote the Horvitz-Thompson estimators of the sums of the characteristic y and the x vector of the auxiliary characteristics, $\hat{\boldsymbol{x}}$ the vector of the sums of the x characteristics known for the population, and $\hat{\boldsymbol{B}}_{\boldsymbol{s}}$ the vector of the regression parameters of y on the x characteristics estimated on the basis of the sample.

The GREG weights can be directly fitted to continuous variables or to their sums. This means for the weighting of the cross-sectional sample that simultaneously the sample can be adjusted to the distribution of establishments and to the distribution of employees.

During the weighting, there are checks of whether the factors in the individual cells are becoming too high or too low, or whether there is no case at all in a cell. In these cases the cell is amalgamated with an appropriate neighbouring cell.

For cross-sectional evaluations, the cross-sectional weighting factor HR2020Q must be used.

# 8 Definition of Longitudinal Cases and Projection

Longitudinal or panel analyses allow researchers to trace developments in individual establishments over a longer period. Due to its large net sample and its long duration of meanwile 28 survey waves, the IAB Establishment Panel offers a wide range of options for such analysis. Table 8 gives an overview of the number of valid interviews for different starting years in order to produce a balanced panel.

**Table 8: Overview of the number of evaluable interviews for different starting years (balanced panel)**

| | Starting year | | | | | | |
|---|---|---|---|---|---|---|---|
| wave | 1996 | 2000 | 2003 | 2007 | 2009 | 2012 | 2016 |
| 1996 | – | – | – | – | – | – | – |
| 1997 | 6,822 | – | – | – | – | – | – |
| 1998 | 5,597 | – | – | – | – | – | – |
| 1999 | 4,654 | – | – | – | – | – | – |
| 2000 | 4,004 | – | – | – | – | – | – |
| 2001 | 3,479 | 10,840 | – | – | – | – | – |
| 2002 | 2,999 | 8,762 | – | – | – | – | – |
| 2003 | 2,610 | 7,295 | – | – | – | – | – |
| 2004 | 2,286 | 6,363 | 12,775 | – | – | – | – |
| 2005 | 2,011 | 5,542 | 10,771 | – | – | – | – |
| 2006 | 1,761 | 4,854 | 9,118 | – | – | – | – |
| 2007 | 1,524 | 4,242 | 7,845 | – | – | – | – |
| 2008 | 1,382 | 3,774 | 6,889 | 12,567 | – | – | – |
| 2009 | 1,233 | 3,338 | 6,037 | 10,598 | – | – | – |
| 2010 | 1,095 | 2,973 | 5,324 | 9,118 | 12,524 | – | – |
| 2011 | 990 | 2,666 | 4,759 | 7,991 | 10,653 | – | – |
| 2012 | 868 | 2,361 | 4,199 | 6,983 | 9,132 | – | – |
| 2013 | 781 | 2,108 | 3,761 | 6,179 | 8,019 | 12,622 | – |
| 2014 | 682 | 1,883 | 3,352 | 5,430 | 6,970 | 10,515 | – |
| 2015 | 603 | 1,674 | 2,951 | 4,800 | 6,116 | 9,010 | – |
| 2016 | 537 | 1,520 | 2,667 | 4,287 | 5,447 | 7,920 | – |
| 2017 | 474 | 1,348 | 2,363 | 3,792 | 4,813 | 6,950 | 12,614 |
| 2018 | 412 | 1,173 | 2,060 | 3,343 | 4,230 | 6,071 | 10,505 |
| 2019 | 337 | 971 | 1,708 | 2,754 | 3,538 | 5,047 | 8,476 |
| 2020 | 261 | 735 | 1,300 | 2,088 | 2,680 | 3,824 | 6,267 |

## 8.1 Panel Case definition for Longitudinal Cases

For longitudinal weighting, all the establishments which were a panel case the previous year and all new establishment numbers of the supplementary and extension sample from the subsequent years are generally considered as panel cases. To count as a panel case, information must be available for every survey year from the respective first interview to the latest survey, in the form either of a valid interview or the information that the establishment (or the establishment number as applicable) has expired. We therefore differentiate between "panel cases with interview" (panel cases for which valid questionnaires are held for every individual year up to the current survey) and "no longer operational panel cases" (in which from the date on which the establishment ceased operations onwards the only information held is that the establishment no longer exists). For the panel cases with interview, it is also necessary to ensure that the information obtained applies to the same establishment unit every year.[25]

The subgroups comprising the respective panel cases can be shown using the longitudinal section for 2012 – 2020:

- All cross-sectional cases from the 2012 wave for which information is available from the 2012 wave onwards.
- Establishments from the supplementary and extension samples 2013 (only "new establishment numbers"), for which information from the 2013 wave onwards is available.
- Establishments from the supplementary and extension samples 2014 (only "new establishment numbers"), for which information from the 2014 wave onwards is available.
- Establishments from the supplementary and extension samples 2015 (only "new establishment numbers"), for which information from the 2015 wave onwards is available.
- Establishments from the supplementary and extension samples 2016 (only "new establishment numbers"), for which information from the 2016 wave onwards is available.
- Establishments from the supplementary and extension samples 2017 (only "new establishment numbers") for which information from the 2017 wave onwards is available.
- Establishments from the supplementary and extension samples 2018 (only "new establishment numbers") for which information from the 2018 wave are available.
- Establishments from the supplementary and extension samples 2019 (only "new establishment numbers") for which information from wave 2019 are available.
- Establishments from the supplementary and extension samples 2020 (only "new establishment numbers") which participated in 2020 wave.

To make it easier for the users of the IAB Establishment Panel to identify different subgroups for cross-sectional and in particular for longitudinal analyses, the relevant subgroups have been marked. For this the following information has to be appropriately combined:

- Field result of the ongoing wave (valid interview with/without employees subject to social insurance contributions as of the reference date, establishment no longer operational, cases that did not respond the previous year but are willing to be surveyed again/final non-responses)

25 In companies with several establishments in particular, it can be the case that the interviewees provide details of different units in different years (e.g. once about the local establishment, once about the entire company). This information (from the address protocol or as the result of the validation and editing process) is taken into account when forming the wave code (WELLwxyz) and is thus available for defining the panel cases.

- Was the establishment surveyed in the previous wave (respondents from the previous wave, non-responses from the previous year that are willing to be surveyed again, supplementary and extension sample)?
- Was the same establishment surveyed as last time?

In consultation with the IAB we have developed the following concept for this. Every case is given a unique identifier in each wave which takes account of the above criteria. This so-called wave code is stored in the variables WELLwxyz, where wxyz stands for the year in which the survey took place (thus WELL1993 for wave 1 in 1993, WELL1994 for wave 2 in 1994 etc.). This labelling takes place using a letter of the alphabet (see Table 9).

**Table 9: Group identifiers in the variable WELLwxyz**

| | | Labelling letter *with* employees subject to social insurance contributions as per the respective reference date[26] | Labelling letter *without* employees subject to social insurance contributions as per the respective reference date[26] |
|---|---|---|---|
| 1. | Cases **with** interview in the ongoing wave | | |
| 1.1 | Cases surveyed for the **first time** (= at the date of drawing) from the basic sample and the respective supplementary or top-up samples[27] | **A** | Not permitted |
| 1.2 | Cases surveyed **repeatedly** | | |
| 1.2.1 | **with** interview the previous year | | |
| 1.2.1.1 | **same** unit interviewed as previous year | **B** | **C** |
| 1.2.1.2 | **different** unit interviewed to previous year | **D** | Not permitted |
| 1.2.2 | **without** interview in previous year[28] | **E** | Not permitted |
| 2. | Cases **without** interview in the ongoing wave | | |
| 2.1 | Non-response that can be surveyed again in future | | **H** |
| 2.2 | Cases from earlier extensions that can no longer be surveyed[29] | | **W** |
| 2.3 | Non-responses that can no longer be surveyed[30] | | **X** |
| 2.4 | No longer operational establishments (according to field result, editing or BA file respectively) | | |
| 2.4.1 | in the ongoing wave | | **Y** |
| 2.4.2 | earlier than this | | **Z** |

In principle, other longitudinal sections can also be defined, however there are no weighting factors available for this.

26 The cut-off date is always June 30 of the previous year, so for wave 28 (2020), for example, June 30, 2019.The weighting takes place using the questionnaire information and the targets from the BA establishment file as of this reference date. No newer information from the BA establishment file is yet available as per the respective weighting date.

27 Note: Dropouts from wave 1 who were interviewed for the first time in wave 2 and dropouts from the basic sample East (wave 4) who were interviewed for the first time in wave 5 do not belong to this group; these cases were assigned the code letter H in wave 1 or wave 4, respectively, and the code letter E in the follow-up wave.

28 Since 2002 non-responses from the previous year (H cases) have been treated in the same way as establishments being surveyed for the first time and the employee details from the previous year according to the BA used for identification. For this reason, there are no longer any G cases (establishments repeatedly surveyed without an interview the previous year and different unit to that surveyed the year before last).

29 Thus e.g. the cases from the extension sample in 1997 in agriculture in Mecklenburg-West Pomerania, or the cases from the Halle Institute for Economic Research (IWH) extension sample in 1998 for the construction sector in East Germany.

30 These include (a) establishments which declined to be interviewed, unless they expressly consented to being interviewed again the following year, and (b) non-responses from the previous wave from which no interview has also been obtained in the ongoing wave (i.e. the combination HH is not permitted and becomes HX).

## 8.2 Overview of the Longitudinal Sections provided with 2020

There are three longitudinal sections for which a panel weighting was undertaken for the 2020 wave. Table 10 depicts the case numbers for the respective longitudinal sections from 2009, 2012 and 2016 onwards.

As already mentioned above, the definition of panel cases essentially comprises all the establishments which were a panel case the previous year – either with an interview or as a no longer operational establishment – and all the establishments that were newly founded between the reference date of the year before last and the reference date of the previous year. For this reason, the number of newly-founded establishments (in the sample) is identical for all longitudinal sections. After the longitudinal section has existed for a certain period of time, the number of newly-founded establishments more or less counterbalances the number of non-responses. Hence after 3 – 5 waves, largely stable case numbers are achieved. For this reason, and because the longitudinal sections shown in Table 10 started with relatively similar case numbers, the number of panel cases in the individual longitudinal sections (with exception of the new longitudinal section of 2016) in 2020 is at a similarly high level.

**Table 10: Overview of all longitudinal sections since 2009**

| Starting year of the respective longit. section | No. of panel cases in start year+1 | No. of panel cases 2020 | Thereof are.. | No. of repeaters | No. of newly founded current wave | No. of no longer operational |
|---|---|---|---|---|---|---|
| 2009 | 14,308 | 9,849 | | 6,143 | 616 | 3,090 |
| 2012 | 14,186 | 9,472 | | 6,719 | 616 | 2,137 |
| 2016 | 14,250 | 9,557 | | 7,972 | 616 | 969 |

## 8.3 Longitudinal Weighting Process

Specific panel weighting factors are calculated for every longitudinal section. As for the cross-sectional weighting, the panel weighting takes place in the form of a projection onto the population. The aim of the panel weighting is that panel analyses using the respective longitudinal weighting factor should as far as possible generate the same distributions for all the waves involved as those from cross-sectional analyses of the individual waves.

Panel weighting essentially takes place in eight steps, each forming the basis for the next, with East and West Germany being differentiated in all steps. Unlike the cross-sectional weighting, for the longitudinal weighting a weighting at federal state level takes place only in the first step. In the subsequent steps the federal state level is no longer monitored. In total the eight weighting steps of the longitudinal weighting are repeated iteratively for as long as this is necessary and purposive.

The first step is to adapt the latest available cross-sectional cases to the structure of the population (number of establishments and employees subject to social insurance contributions as of 30 June of the previous year). The entry factor is the previous year's panel weighting factor and for new establishment numbers the cross- sectional weighting factor of the current wave.

The next stage is to adjust the key figure for the stock of establishments for each individual year included in the longitudinal section, and for the establishment numbers that in the meantime have become no longer operational or are new. Each year's stock includes the

establishments with at least one employee subject to social insurance contributions as of the reference date. The new establishment numbers consist of those from the stock that had no employees subject to social insurance contributions as of the reference date one year before. The no longer operational establishment numbers are those in the stock that no longer had any employees subject to social insurance contributions as of the reference date one year later. This step depicts the recovery or termination of establishment numbers over the course of time.

The third step is to adjust for the establishments that are surviving or no longer operational from the different entry cohorts.

The fourth and fifth steps take account of the establishment dynamics (growth and shrinkage of establishments). In particular in the longitudinal sections that have been running for a longer period, an extremely large number of combinations is possible in respect of the development of the number of employees. Simply because of the limited case numbers, it is not possible to obtain a finely differentiated picture of the possible development paths (changes between individual establishment size classes from one year to the next year). We have therefore applied a simplified procedure for the panel weighting in consultation with the IAB: a change of establishment size class is only taken into account between the starting wave of the respective longitudinal section (or as applicable for new establishment numbers, the first time they were surveyed) and the latest available data. No check is kept on changes occurring in between. Due to extremely low numbers of cases in certain combinations, the theoretically possible combinations are compounded as follows: as of the latest available data the establishment is in the same establishment size class as it was at the beginning, or has grown, or has shrunk. It is self-evident that when this approach is used the development paths of establishments can only be represented approximately during the panel weighting. In the fourth step, the establishment dynamics are taken into account for the establishments from the stock of the starting wave of the respective longitudinal section. In the fifth step, this is done for the “new” establishment numbers added during the longitudinal period.

The sixth step is to correct any disproportional non-responses depending on the answers to individual questions from the previous year’s survey. This is done by undertaking multivariate analyses of non-responses in comparison between the previous wave and the most recent available data.

The two subsequent steps involve adapting the cases belonging to the respective cross-section to the requirements of the stratification matrix for every survey year included in the longitudinal section. To limit the number of weighting cells, only the establishment size classes (seventh step) and a simplified sector structure with six categories (step eight) are taken into account (cf. Appendix Table 13).

In individual cases using the aforementioned parameters during the panel weighting can lead to extremely high weighting factors. To limit the resulting problems of outliers, only weighting factors up to a maximum of 4,000 are permitted for the panel weighting. As a consequence, the theoretically necessary target numbers for individual characteristic values may not be reached due to the weighting process. Thus, for example, new establishment numbers in certain years are substantially underestimated during the panel weighting.

For longitudinal evaluations, the following factors are to be used:

- For the longitudinal section 2009 - 2020: HR09_20P
- For the longitudinal section 2012 - 2020: HR12_20P
- For the longitudinal section 2016 - 2020: HR16_20P

# 9 General Information about Evaluation

- Descriptive results should always be calculated and interpreted on a projected level, so that the disproportional sample structure and selective non-responses are corrected.
- The results should always be interpreted against the background of the underlying (unweighted) case numbers. The lower the number of cases considered, the more uncertain the results. The area of uncertainty for different unweighted case numbers is shown in the error tolerance table (cf. Appendix). The case number of 15,000 establishments can exhibit differences of two percentage points and more as significant differences.
- Even when the projection factor contains an adjustment proportional to the numbers of employees, we recommend always showing employee numbers rounded to full thousands.
- The panel weighting does indeed take place in the form of a projection, however for methodological reasons it is associated with greater inaccuracies in terms of the adjustment to the different target structures than the respective cross-sectional weightings, because the target structures of the various cross-sections, which fluctuate to some extent, has to be achieved with just one projection factor. This limitation applies to both the projected number of establishments and – to an even greater extent – to projected employee figures from the IAB Establishment panel.
- We therefore urgently recommend that when analyses take place using longitudinal factors not to show absolute figures. Projected absolute figures (even if rounded to full thousands) give the reader the impression of an accuracy that cannot be achieved with the IAB Establishment Panel when longitudinal analyses are used.
- As with all sample-based data, we recommend to use distribution measures, in other words statements of the type “X % of establishments have grown since 2003, Y % have contracted” or “The number of employees by Year Y developed better/worse in establishments which invested more than average in Year X than was the case for establishments which made no investments in Year X”. It should also be borne in mind that the percentage values obtained can also be afflicted with inaccuracies, so the values should if anything be interpreted as orders of magnitude.
- When undertaking analyses, account should be taken of the type of survey (conducted completely face-to-face through to entirely self-completed by the respondent) in the form of a third variable control.[31] In the 2020 wave this information is stored in the variable befart20.
- When undertaking analyses, particularly in comparisons across time, account needs to be taken of changes to the delineations of the sectors, the regional classification and the

31 With the introduction of the mail survey method in 2000 “a test was carried out as to what extent distortions of content derived from the mix of methods. Differences in answering behaviour that were to some extent significant became apparent between the face-to-face and mail interviews. Account should therefore be taken of the characteristic of the survey method when analysing the federal states concerned.” (Fischer et al. 2008: 14).

population. Such changes on their own can lead to sometimes substantial changes in the parameters and distributions. Thus when the employee statistics were revised with the 2015 wave, new groups of persons were included; hence the increase at that time in the number of employees subject to social insurance contributions by a good three percent can also be ascribed to the revision of the employee statistics, and not just to changes in the real employee situation.

Moreover, in 2009 the last change of the stratification matrix was made in the course of the changeover from the WZ2003 to the WZ2008 classification of economic sectors. Changes in the shares of establishments by sector can therefore only be attributed to real changes in the economy structure to a limited extent.

Since 2007 East and West Berlin have been aggregated. In the consequence of this, Berlin was excluded from the extention sample in manufacturing industry in eastern Germany at the request of the Leibniz Institute for Economic Research in Halle (IWH). Since then, this extention sample only includes eastern German territorial states. The IWH sample since 2007 is therefore only partially comparable with the IWH sample before 2007.

- Generally when undertaking time series and panel analyses, account should be taken of changes to the question or the individual items, so that differences in results are actually ascribable to real changes and not to changes to the question and/or the answer requirements. Account should also be taken of the respective timeframe. Thus establishment sizes refer to the reference date 30 June, flow figures to the first half year, and some questions – e. g. the business volume –relate to the last year or the financial year that has recently expired.

# 10 Bibliography


Cochran, William Gemmell (1968): The effectiveness of adjustment by subclassification in removing bias in observational studies. Biometrics. 24. 295-313.

Deming, W. Edwards and Stephan, Frederick F. (1940): On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. Annals of Mathematical Statistics. 11. 427-444.

Destatis (2020): https://www.destatis.de/DE/Themen/Wirtschaft/Konjunkturindikatoren/Arbeitsmarkt/arb510a.html. Retrieved on 10.06.2021.

Deville, J.-C., Särndal, C.-E. and Sautory, O., 1993: Generalized Raking Procedures in Survey Sampling, Journal of the American Statistical Association, Vol. 88, No. 423, pp.1013-1020)

Ellguth, Peter; Kohaut, Susanne, and Iris Möller (2014): The IAB Establishment Panel – methodological essentials and data quality. In: Journal of Labour Market Research 47. 27 – 41.

Fischer, Gabriel, and Janik, Florian; Müller, Dana, and Alexandra Schmucker (2008): Das IAB Establishment Panel– von der Stichprobe über die Erhebung bis zur Hochrechnung. In: FDZ-Methodenreport 01/2008. Bundesagentur für Arbeit: Nürnberg.

Fischer, Gabriel; Janik, Florian; Müller, Dana, and Alexandra Schmucker (2009): The IAB Establishment Panel – Things Users Should Know. In: Schmollers Jahrbuch 129. 133 – 148.

Frank, Thomas, and Christopher Grimm (2015): Beschäftigungsstatistik Revision 2014. Bundesagentur für Arbeit: Nürnberg.

Fritsch, Michael, and Udo Brixy (2004): The Establishment File of the German Social Insurance Statistics. In: Schmollers Jahrbuch 124. 183 – 190.

Janik, Florian, and Susanne Kohaut (2009): Why Don't They Answer? – Unit Non-Response in the IAB Establishment Panel. In: FDZ-Methodenreport 07/2009. Bundesagentur für Arbeit: Nürnberg.

Krosnick, John A., Sowmya Narayan, and Wendy R. Smith (1996): Satisficing in Surveys: Initial Evidence. In: Advances in Survey Research. San Francisco: Jossey-Bass. 29 – 44

Sommer, Rudolf; Unholzer, Gerhard, and Erich Wiegand (1999): Standards zur Qualitätssicherung in der Markt- und Sozialforschung. ADM: Frankfurt a. M.

Winters, Jutta, and Andrea Kargus (2012): Die ProIAB: Stets zur Stelle im Dienste der Wissenschaft. Ein Gespräch mit Christine Hense und Christian Sprenger. In: IAB-Forum, No. 2. 34 – 39.

# 11 Appendix

The study “Employment Trends – Employer Survey 2020” (IAB Establishment Panel 2020) was undertaken on behalf of the Institute for Employment Research (IAB) of the Federal Employment Agency (BA).

Regional and sectoral expansions of the sample were commissioned by:

- The Senate Department for Integration, Labour and Social Affairs of the Federal State of Berlin
- The Ministry for Labour, Social Affairs, Health, Women and Family in the Federal State of Brandenburg
- The Ministry for Economic Affairs, Labour and Health for the Federal State of Mecklenburg-West Pomerania
- The Saxony State Ministry for Economic Affairs, Labour and Transport
- The Ministry of Labour, Social Affairs and Integration Saxony-Anhalt
- The Thuringian Ministry for Labour, Social Welfare, Health, Women and Family Affairs
- The Halle Institute for Economic Research (IWH)
- The Ministry of Finance and Economic Affairs Baden-Württemberg
- The Bavarian Ministry of Labour and Social Affairs, Family and Integration
- The Senator for Economic Affairs, Labour and Ports of the Federal State of Bremen
- The Hessian Ministry of Economics, Energy, Transport and Regional Development
- The Ministry of Economic Affairs, Employment,Transport and Digitalization of Lower Saxony
- The Ministry of Labour, Health and Social Affairs of North Rhine-Westphalia
- The Ministry for Social Affairs, Labour, Health and Demography of Rhineland-Palatinate
- The State Ministry of Saarland for Economic Affairs, Labour, Energy and Traffic

## Table 11: Classification of economic activities by 19 sectors for sampling and cross-sectional weighting from the 2010 wave onwards

| Variable BR19BAxy | Sector | WZ2008 code | Questionnaire sector |
|---|---|---|---|
| 1 | Agriculture, forestry and fishing | 1 – 3 | 1 |
| 2 | Mining and quarrying, electricity, gas and water supply; sewerage and waste management | 05 – 09, 35 – 39 | 2 – 3 |
| 3 | Food products, beverages and tobacco | 10 – 12 | 4 |
| 4 | Consumer products (excluding manufacture of timber products) | 13 – 18 | 5 – 6 |
| 5 | Industrial goods (including manufacture of timber products) | 19 – 24 | 7 – 10 |
| 6 | Capital and consumer goods | 25 – 33 | 11 – 17 |
| 7 | Construction | 41 – 43 | 18 – 19 |
| 8 | Wholesale, sale and repair of motor vehicles | 45 – 46 | 20 – 21 |
| 9 | Retail | 47 | 22 |
| 10 | Transport and warehousing | 49 – 53 | 23 |
| 11 | Information and communication | 58 – 63 | 24 |
| 12 | Hotels and restaurants | 55 – 56 | 25 |
| 13 | Financial and insurance services | 64 – 66 | 26 |
| 14 | Economic, scientific and freelance services | 68 – 82 | 27 – 36 |
| 15 | Education | 85 | 37 |
| 16 | Health and social services | 86 – 88 | 38 |
| 17 | Other services | 90 – 93, 95, 96 | 39 – 41 |
| 18 | Representations of interests | 94 | 42 |
| 19 | Public administration, defence, social security | 84 | 43 |

## Table 12: Establishment size class classifications for sampling and cross-sectional and longitudinal weightings

| Employees subject to social insurance contributions on 30 June of the previous year |
|---|
| 1 – 4 |
| 5 – 9 |
| 10 – 19 |
| 20 – 49 |
| 50– 99 |
| 100 – 199 |
| 200 – 499 |
| 500 – 999 |
| 1000+ |

## Table 13: Classification of economic activities by 6 sectors for longitudinal weighting from the 2009 wave onwards

| Sector | Code from classification by 19 sectors |
|---|---|
| Agriculture and forestry, fishing | 1 |
| Manufacturing industry | 3 – 6 |
| Other production industry | 2, 7 |
| Retail/transport and warehousing/hotels and restaurants | 8 – 12 |
| Financial and insurance services/business services | 13, 14 |
| Public and private services | 15 – 19 |

# Imprint

**FDZ-Methodenreport 7|2021 EN**

**Date of publication**
15 December 2021

**Publisher**
Research Data Centre (FDZ)
of the Federal Employment Agency (BA)
in the Institute for Employment Research (IAB)
Regensburger Str. 104
D-90478 Nuremberg



**Download**
https://doku.iab.de/fdz/reporte/2021/MR_07-21_EN.pdf

**Documentation version**
IABBP9320_EN_v1_dok1, DOI: 10.5164/IAB.FDZM.2107.en.v1

**Dataset version**
IABBP 9320 v1, DOI: 10.5164/IAB.IABBP9320.de.en.v1

**All publications in the series "FDZ-Methodenreport" can be downloaded from**
https://fdz.iab.de/en/FDZ_Publications/FDZ_Publication_Series/FDZ-Methodenreporte.aspx

**Website**
http://fdz.iab.de

---

**Corresponding author**
Susanne Kohaut
Phone: +49 911 179-3253
Email: susanne.kohaut@iab.de