



FORSCHUNGSDATENZENTRUM
der Bundesagentur für Arbeit im Institut für
Arbeitsmarkt- und Berufsforschung

FDZ-DATENREPORT

Documentation of labour market data

13|2021 EN Panel “Establishments in the Covid-19 Crisis” – 20/21

A longitudinal study in German establishments – waves 1 - 14

Nils Backhaus, Lutz Bellmann, Patrick Gleiser, Sophie Hensgen, Christian Kagerl, Theresa Koch, Corinna König, Eva Kleifgen, Ute Leber, Moritz Kuhn, Michael Moritz, Laura Pohlan, Swantje Robelski, Duncan Roth, Malte Schierholz, Sabine Sommer, Jens Stegmaier, Anita Tisch, Matthias Umkehrer, Armin Aminian



Bundesagentur für Arbeit

Panel “Establishments in the Covid-19 Crisis” – 20/21

A longitudinal study in German establishments – waves 1 - 14

Nils Backhaus, Lutz Bellmann, Patrick Gleiser, Sophie Hensgen, Christian Kagerl, Theresa Koch, Corinna König, Eva Kleifgen, Moritz Kuhn, Ute Leber, Michael Moritz, Laura Pohlan, Swantje Robelski, Duncan Roth, Malte Schierholz, Sabine Sommer, Jens Stegmaier, Anita Tisch, Matthias Umkehrer, Armin Aminian

Die FDZ-Datenreporte beschreiben die Daten des FDZ im Detail. Diese Reihe hat somit eine doppelte Funktion: zum einen stellen Nutzerinnen und Nutzer fest, ob die angebotenen Daten für das Forschungsvorhaben geeignet sind, zum anderen dienen sie zur Vorbereitung der Auswertungen.

FDZ-Datenreporte (FDZ data reports) describe FDZ data in detail. As a result, this series of reports has a dual function: on the one hand, those using the reports can ascertain whether the data offered is suitable for their research task; on the other, the data can be used to prepare evaluations.

Contents

1	Introduction	5
2	Data description	7
2.1	Outline.....	7
2.2	Volume structure.....	9
3	Data collection	10
3.1	Contents and method.....	10
3.2	Population and sampling	12
4	Data preparation	15
4.1	Data organisation.....	15
4.2	Data validation and correction.....	16
4.3	Weighting	18
4.4	Linkage with administrative establishment data (BHP).....	19
4.5	Anonymisation	20
5	Information for data use.....	21
5.1	Preparations for data analysis.....	21
5.2	Stata evaluation example.....	21
5.3	Data access and additional material.....	23

List of tables

Table 1: Content-related characteristics	7
Table 2: Methodological characteristics	8
Table 3: Data access	8
Table 4: Volume structure	9
Table 5: Overview of topics in the flexible module.....	11
Table 6: Stratification variables of the gross samples.....	13
Table 7: Variables carried forward	16
Table 8: Classification of missing values.....	17

Zusammenfassung

Dieser Datenreport beschreibt die Daten der im Rahmen des Projekts „Betriebe in der Covid-19-Krise“ durchgeführten Betriebsbefragung. In regelmäßigen Abständen ist dabei eine repräsentative Stichprobe von 1.500 bis 2.000 Betrieben aus verschiedenen Wirtschaftszweigen und Betriebsgrößen befragt worden.

Durch ein hochfrequentes, rotierendes Panel-Design bietet die Befragung eine Datenbasis, welche die vielseitigen Auswirkungen der Pandemie auf Betriebe in Deutschland in einem zeitlich engen Takt erfasst. Bislang sind vierzehn Wellen (Zeitraum: 03.08.2020 – 18.06.2021) für externe Forscherinnen und Forscher verfügbar.

Details zu verschiedenen Zugangswegen, den Voraussetzungen der Nutzung sowie zur Antragstellung des Datensatzes können der FDZ-Internetseite unter <http://fdz.iab.de/> entnommen werden.

Abstract

This data report describes the data from an establishment survey conducted as part of the "Establishments in the Covid-19 Crisis" project. A representative sample of 1,500 to 2,000 establishments of different sizes and from various economic sectors was surveyed at regular intervals.

By utilising a high-frequency, rotating panel design, the survey provides a database that closely tracks the consequences of the pandemic. So far, fourteen waves (period: 08/03/2020 - 06/18/2021) are available to external researchers.

Details on data access, conditions of use and how to request the dataset can be found on the FDZ website at <http://fdz.iab.de/>.

Keywords

Establishment survey, panel dataset, data documentation, Covid-19 pandemic

Acknowledgements

We wish to thank our cooperation partners, in particular the Federal Institute for Occupational Safety and Health (Bundesanstalt für Arbeitsschutz und Arbeitsmedizin (BAuA)), and the Cluster of Excellence ECONtribute of the Universities of Cologne and Bonn, for their participation in the study in terms of funding as well as content and conception, and Kantar Public for implementing the project and conducting the survey.

1 Introduction

The Covid-19 crisis poses major challenges to establishments. In order to contain the virus, the authorities initiate restrictions and temporary lockdowns for large areas of social life. As a result, many establishments are obliged to reduce their business activity at times or to change their work processes in many ways in order to meet the dynamic requirements of occupational health and safety. In addition, the crisis causes supply shortages as well as a drop in demand for certain goods and services. However, establishments are affected by these challenges to very different degrees. Many are only marginally affected by the crisis, while others have to suspend their business activity to a large extent. Furthermore, establishments have different concepts and strategies at their disposal to satisfy the official restrictions, for example, expanding the use of homeworking or implementing comprehensive hygiene concepts. The panel “Establishments in the Covid-19 Crisis” examines the economic and operational impact of the pandemic on establishments in Germany.

Utilising a high-frequency, rotating panel design, establishments of different sizes and from various economic sectors of the economy are interviewed by telephone in a three- to four-week cycle regarding the impact of the Covid-19 crisis on various areas. The survey results in a panel consisting of 20 planned waves that cover a period starting on 3 August 2020 and presumably lasting until the end of 2021. Between 1,500 and 2,000 establishments (first-time and panel respondents) are surveyed in each wave.

The data obtained in the establishment survey permit representative cross-sectional and longitudinal analyses of the diverse effects that the Covid-19 crisis is having on establishments in Germany. The short intervals between survey waves also make it possible to closely track all phases of the pandemic. In order to increase the analysis potential even further, the data can be linked with administrative data gathered by the Federal Employment Agency (Bundesagentur für Arbeit).

The datasets originated in the project “Establishments in the Covid-19 Crisis”. Parts of the project were conducted in the context of a research collaboration with the Federal Institute for Occupational Safety and Health (Bundesanstalt für Arbeitsschutz und Arbeitsmedizin (BAuA)) and the Cluster of Excellence ECONtribute of the Universities of Cologne and Bonn. In three of the waves carried out so far, the BAuA was involved in the content-related conception of the survey instrument as well as the funding. The Cluster of Excellence contributed financially and conceptually to wave 13 in particular.

This data report documents the data gathered in the first fourteen waves of the establishment survey (period: 3 August 2020 to 18 June 2021). It contains information about the survey contents, the survey method, sampling, data preparation, anonymisation, weighting and data organisation. Questionnaires and an overview of the variables are available separately from the Research Data Centre of the Federal Employment Agency. Further waves of the establishment survey will be published by the Research Data Centre at regular intervals until the end of the study (presumably December 2021).

External researchers can access the datasets via the Research Data Centre. Various data access options can be found on the official website (<http://fdz.iab.de/>).

2 Data description

2.1 Outline

Table 1: Content-related characteristics

Category	Description
Current data version	BeCovid 2020/2021
Topics/variable groups	Surveyed in every wave: Establishment structure Workforce development Liquidity reserves Extent and type of economic effects due to the Covid-19 crisis Short-time work (from wave 3 onwards) Surveyed in individual waves: Occupational safety and health Vocational training and training bonus Homeworking Further training Exports and imports Effects of the lockdown Digitalisation Atypical employment Work organization Company measures for mental health
Research unit	Establishments with at least one employee subject to social security as of the reference date. Excluded economic sectors: “public administration, defence, social security” (WZ 2008 (Classification of Economic Activities 2008):84), “private households” (WZ 2008: 97+98) and “extraterritorial organisations” (WZ 2008: 99). Reference date for waves 1 to 4: 30 November 2019 Reference date from wave 5 onwards: 28 February 2020
Number of cases	1,500 to 2,000 establishments per wave (first-time and repeat respondents)
Period	3 August 2020 to 18 December 2020 (three-week rhythm) 18 January 2021 to 26 May 2021 (three-week rhythm) 7 June 2021 to 18. June 2021 (four-week rhythm)
Time reference	Survey date
Regional structure	East/west
Territorial allocation	As of respective reference date

Table 2: Methodological characteristics

Category	Description
Gross samples	<p>Repeat-respondent sample: Establishments:</p> <ul style="list-style-type: none"> • which agreed to participate again in the previous interview. • which drop out temporarily between individual waves. • which were successfully surveyed less than six times. • whose first interview was no longer than seven waves ago. • whose last interview was no longer than four waves ago. <p>Refreshment sample: Establishments which:</p> <ul style="list-style-type: none"> • are not participating in the IAB Job Vacancy Survey or the IAB Establishment Panel at the time of the survey. • have not yet participated in the survey or had not been contacted previously. • did not refuse to participate in the survey after being contacted the first time.
Data collection method	Telephone interviews (CATI)
Institutions involved in the survey	Institute for Employment Research (IAB), Federal Institute for Occupational Safety and Health (BAuA), Kantar Public
File format and size	STATA (approx. 500 KB per wave)
File architecture	By wave

Table 3: Data access

Category	Description
Modes of access	On-site access followed by remote data access
Degree of anonymisation	Weakly anonymous
Sensitive variables	<p>Survey data: none</p> <p>Administrative establishment data: place of work – district (Kreis) (ao_kreis), subclass of economic activity (w08_5), first appearance of establishment number (grd_dat), last appearance of establishment number (lzt_dat)</p>
Details for citation and data documentation	<p>Data: " Backhaus, Nils; Bellmann, Lutz; Gleiser, Patrick; Hensgen, Sophie; Kagerl, Christian; Koch, Theresa; König, Corinna; Kleifgen, Eva; Kuhn, Moritz; Leber, Ute; Moritz, Michael; Pohlan, Laura; Robelski, Swantje; Roth, Duncan; Schierholz, Malte; Sommer, Sabine; Stegmaier, Jens; Tisch, Anita; Umkehrer, Matthias; Aminian, Armin (2021): "Panel „Establishments in the Covid-19 Crisis" (IAB BeCovid) – Welle 01-14". Research Data Centre of the Federal Employment Agency (BA) at the Institute for Employment Research (IAB). DOI: 10.5164/IAB.BeCovid_W01-W14.de.en.v1</p> <p>The data access was provided via on-site use at the Research Data Centre (FDZ) of the German Federal Employment Agency (BA) at the Institute for Employment Research (IAB) and/or remote data execution."</p> <p>Data documentation: Nils Backhaus, Lutz Bellmann, Patrick Gleiser, Sophie Hensgen, Christian Kagerl, Theresa Koch, Corinna König, Eva Kleifgen, Moritz Kuhn, Ute Leber, Michael Moritz, Laura Pohlan, Swantje Robelski, Duncan Roth, Malte Schierholz, Sabine Sommer, Jens Stegmaier, Anita Tisch, Matthias Umkehrer, Armin Aminian (2021): Panel on "Establishments in the Covid-19 Crisis – 20/21". A longitudinal study in German establishments – wave 1 - 14. FDZ-Datenreport, 13/2021 (en), Nuremberg. DOI: 10.5164/IAB.FDZD.2113.en.v1</p>

2.2 Volume structure

In each wave the establishment survey consists of first-time- and repeat respondents. Table 4 shows the number of observations and the composition of the available datasets. A minimum of 1,500 establishments were recorded in each wave. After a few waves, the proportion of first-time respondents fell to an envisaged level of approximately 30-40 percent and serves among other things as a refreshment sample. In wave 8 the proportion of first-time respondents increased once again to 54 percent as a large proportion of the panel establishments were rotated out of the sample.

Table 4: Volume structure

Dataset	Week number	Year	Size of net sample	Proportion of first-time respondents	Proportion of repeat respondents
BeCovid_W1_v2.dta	32, 33	2020	1,723	100%	0%
BeCovid_W2_v2.dta	35, 36	2020	1,556	53.92%	46.08%
BeCovid_W3_v2.dta	38, 39	2020	1,733	40.33%	59.67%
BeCovid_W4_v2.dta	41, 42	2020	1,794	38.97%	61.03%
BeCovid_W5_v2.dta	44, 45	2020	1,760	31.38%	68.62%
BeCovid_W6_v2.dta	47, 48	2020	2,001	38.98%	61.02%
BeCovid_W7_v2.dta	50, 51	2020	1,511	33.09%	66.91%
BeCovid_W8_v2.dta	3, 4	2021	1,999	54.23%	45.77%
BeCovid_W9_v2.dta	6, 7	2021	2,000	34.45%	65.55%
BeCovid_W10_v1.dta	9, 10, 11	2021	2,002	33.72%	66.28%
BeCovid_W11_v1.dta	12, 13, 14	2021	2,009	37.33%	62.67%
BeCovid_W12_v1.dta	16, 17	2021	2,001	30.03%	69.97%
BeCovid_W13_v1.dta	19, 20, 21	2021	2,001	35.52%	64.48%
BeCovid_W14_v1.dta	23, 24	2021	2,001	39.78%	60.22%

3 Data collection

3.1 Contents and method

The content-related and methodological conception of the panel survey “Establishments in the Covid-19 Crisis” is elaborated jointly by the responsible project team at the IAB and various collaboration partners, in particular at the BAuA. The survey is conducted by Kantar Public. It focuses on the economic and structural effects of the Covid-19 crisis on establishments in Germany, as well as on changes in those effects over the course of the pandemic. A total of fourteen individual waves have so far been conducted in the period between 3 August 2020 and 18 June 2021. Table 5 contains information about the starting date and duration of the respective fieldwork periods. New waves were generally conducted at three- to fourweek intervals. The active fieldwork period for each individual wave lasted two weeks. Interviews were conducted solely by telephone (CATI).

The survey is conducted on the basis of a rotating panel design. Establishments are rotated out of the panel after participating a maximum of six times or at the latest seven waves after their first interview. It is irrelevant how often the establishment was actually interviewed. Thus, for example, establishments that were only successfully interviewed twice within the period of seven waves after their first interview also rotate out of the panel. Furthermore, establishments that are particularly especially difficult to contact and could not be interviewed in four consecutive waves are replaced by other establishments. Establishments that are lost as a result of the rotation or non-response are replaced by newly recruited establishments.

Participation in the survey is voluntary at all times. Establishments selected for the survey are sent an invitation letter some weeks beforehand, announcing the upcoming phone call from an interviewer. The letter also contains information about the objectives of the study and about data protection.

The questionnaire contains three types of questions: “structural questions”, “panel questions” and “module questions”. *Structural questions* are usually asked only once, i.e. the first time that an establishment takes part. They collect information regarding basic characteristics of the establishment’s organisation, thereby permitting a differentiated data analysis for different sub-populations. The characteristics include, for example, details about the number of employees before the start of the pandemic, particular aspects of training activity and the organisational structure. Most of the structural details are characteristics that are assumed to remain constant over time. They are therefore not asked of repeat-interviewed establishments again, but are automatically carried forward using the information from the initial survey. More detailed information about the process of carrying information forward can be found in Section [4.1](#). For establishments that have consented to their details being linked with administrative data held by the Federal Employment Agency (92.85 percent of the establishments surveyed), a more extensive range of information concerning the establishment structure is additionally available from the data of the Establishment History

Panel (BHP). This information can be merged with the panel using the establishment identifier “betnr” (see Section 4.4).

Panel questions are asked at regular intervals of both first-time and repeat respondents and frequently refer to the effects of the Covid-19 crisis. They are divided into the topic areas of “extent of impact and economic effects”, “workforce development” and “short-time work”. As the interviews are conducted at very short intervals, this information permits detailed time-series analyses of some relevant indicators pertaining to the Covid-19 crisis. An overview of when individual panel questions were asked is provided in the list of variables that is available separately via the website of the Research Data Centre:

(https://fdz.iab.de/de/FDZ_Establishment_Data/iab_becovid/IABBeCovid_W01-W014.aspx).

The *module questions* are included in a flexible section of the questionnaire whose content changes in each wave. The module questions of one wave generally refer to specific topics that are separate from the rest of the survey. Table 5 provides an overview of the topics covered, as well as information about when the questions were asked. The content of the set of questions can be found in the questionnaire of the respective wave¹.

Table 5: Overview of topics in the flexible module

Topic	Wave	Fieldwork period	Institutions involved
Short-time work, extent of impact	1	03.08 – 18.08.2020	-
Occupational safety and health	2	25.08 – 04.09.2020	BAuA
Vocational training, training bonus	3	14.09 - 25.09.2020	-
Homeworking	4	05.10 – 19.10.2020	BAuA
Further vocational training	5	26.10 – 09.11.2020	RU: Active Labour Market Policies and Integration (IAB)
Future trends, exports, lockdown	6	16.11 – 27.11.2020	RU: Migration and International Labour Studies (IAB)
2021/22 training year, training bonus (II), recruitment	7	07.12 – 18.12.2020	Federal Employment Agency
Short-time work allowance, lockdown, minimum wage, homeworking	8	18.01 – 29.01.2021	Minimum Wage Commission
Digitalisation	9	08.02 – 19.02.2021	Heidelberg University / LISER
Training bonus III, homeworking	10	01.03 – 12.03.2021	Research Data Center (IAB)
Vignette atypical employment. (Hiring behaviour), Corona tests, homeworking	11	22.03. – 07.04.2021	RU: education, qualification and employment (IAB) FB: Businesses and employment (IAB)
Leadership/Performance appraisal, Tests, homeworking, Short-time work allowance	12	19.04. – 30.04.2021	Heidelberg University
Work organisation	13	10.05. – 26.05.2021	Rheinische Friedrich-Wilhelms University Bonn

¹ The questionnaires are available via the FDZ website (https://fdz.iab.de/de/FDZ_Establishment_Data/iab_becovid/IABBeCovid_W01-W14.aspx).

Topic	Wave	Fieldwork period	Institutions involved
Company measures for mental health (+ homeworking)	14	07.05 – 18.06.2021	BAuA

3.2 Population and sampling

The population of the panel comprises all establishments with at least one employee subject to social security. An establishment is defined on the basis of the establishment number concept of the Federal Employment Agency (Bundesagentur für Arbeit 2013). According to that, an establishment denotes a regionally and economically separate unit with employees that is assigned its own establishment number according to certain rules in connection with the social security notification procedure. The following applies in this context²:

- Subsidiaries and branch offices belonging to one employer but located in different municipalities are generally assigned an establishment number of their own.
- Subsidiaries and branch offices belonging to one employer within one municipality are combined to one establishment with one establishment number if they perform the same economic activity.
- Subsidiaries and branch offices belonging to one employer within one municipality are assigned an establishment number of their own if they perform different economic activities.

The sampling is based on the Establishment File of the Federal Employment Agency as of the respective reference date. It contains all establishments with an establishment number. For practical research reasons, the following economic activities were excluded from the sampling procedure:

- public administration, defence; social security (WZ 2008 (Classification of Economic Activities 2008): 84)
- private households (WZ 2008: 97 + 98)
- extraterritorial organisations and bodies (WZ 2008: 99)
- establishments with missing details for the number of employees subject to social security or for the economic activity (information necessary for sampling procedure)

From this population a sample was drawn that was stratified according to establishment size and economic activity. The categories of the stratification variables were selected in such a way as to ensure that sufficient establishments are available per stratum throughout the entire period of the study. Table 6 describes the stratification variables. The stratification corresponds to a cross-tabulation matrix of the two variables, whereby establishments with more than 250 employees were not additionally split according to economic activity. This results in a total of 16 strata. The size of the individual stratum is the result of an employee-optimal allocation strategy (exception: the number of large establishments with more than

² A detailed description of the establishment number logic can be found in Fritsch/Brixy 2004 and on the website of the Federal Employment Agency (https://www.arbeitsagentur.de/datei/dok_ba015200.pdf)

250 employees is too small to carry out this strategy). The sample is therefore drawn disproportionately on the basis of the establishment size. The larger an establishment is, the more likely it is to be included in the sample. This disproportionality is subsequently corrected in the weighting procedure.

The sampling for the first four waves of the panel is based on the establishment file as of the reference date of 30 November 2019. The population described above comprises a total of 2,111,169 establishments. Establishments that are no longer allowed to be contacted for IAB surveys due to a written refusal or are already participating in another IAB establishment survey (N = 30551) are additionally excluded from the sampling process. From the fifth wave onwards the sample is based on the establishment file as of the reference date of 28 February 2020, in which the population consists of 2,096,633 establishments. At the time of sampling for wave five, 2,055,409 of these establishments were available.

Because of the long overall fieldwork period, the sample was divided into tranches. These were drawn at different points in time and transmitted successively to Kantar. There, a telephone number search was conducted for all the establishments. Telephone numbers were successfully identified for a total of 83.49 percent of the sampled establishments. After that, both duplicate phone numbers (2.34 percent) and numbers that were on a contact blacklist at the survey institute (6.65 percent) were eliminated. Furthermore, a screening question was used in the first interview to remove establishments belonging to the public service (0.014 percent). All in all it was possible to contact approximately 75.41 percent of the sampled establishments for an interview.

Due to the panel character, individual waves consist of two sub-samples:

- **Repeat-participant sample:** Comprises all establishments that agreed in their last valid interview to participate again. Establishments that have already taken part in the survey six times as well as those whose first interview was already seven waves ago or whose last valid interview was more than four waves ago are excluded from this rule, as those rotate out of the panel.
- **Refreshment sample:** Consists solely of establishments that have not yet been surveyed at any time during the study, irrespective of how many attempts were previously made to contact the establishment. A minimum of 500 first-time respondents per wave is envisaged.

Table 6: Stratification variables of the gross samples

Variable	Categories
Establishment size	1-9, 10-49, 50-249, 250+

Variable	Categories
(according to number of employees)	
Economic activity (according to Classification of Economic Activities 2008 (Klassifikation der Wirtschaftszweige WZ 2008))	Mining and quarrying; electricity, gas and water supply; construction (codes in WZ 2008: 05-09,35-39,41-43) Manufacturing; agriculture, forestry and fishing (codes WZ 2008: 01-03,10-33) Wholesale and retail trade; transportation; information and communication (codes in WZ 2008: 45-47,49-53,58-63) Accommodation and food services; financial and insurance activities; business-related services; education and other service activities (codes in WZ 2008: 55-56,64-66,68-82,85,90-93,95-96) Health and social work; non-profit organisations; public administration (codes in WZ 2008: 86-88,94)

4 Data preparation

4.1 Data organisation

The panel contains structural information for all the establishments that have taken part in the survey at least once. Some of these details are asked only of first-time respondents in each wave. For repeat respondents, in contrast, this information is always carried forward using the information from the initial survey. A list of all the variables that are carried forward is shown in Table 7. Variables that are carried forward can also be identified by a marker in the variable name (see below).

The variable names consist of a two-part code. The first part consists of a leading chain of letters, which assigns the variable to one of the content-related topic groups. For example, all the variables on the topic of “workforce development” are given the prefix “pe”. The second part consists of a five-digit numerical code. The first two digits simply serve as a counter to distinguish between the different variables belonging to one topic group. The third digit indicates whether it is a variable that is carried forward from previous waves (1) or one that is surveyed in the current wave (0). The last two digits serve as a version number, starting with 0. If the survey method for a variable changes between the individual waves, for example as a result of changes in the wording of the question or the filter procedure, this digit increases by 1.

EXAMPLE: WA04001

WA = ECONOMIC EFFECTS (WIRTSCHAFTLICHE AUSWIRKUNGEN)

04 = QUESTION 04 IN THIS TOPIC AREA

0 = NOT CARRIED FORWARD

01 = THE SURVEY METHOD WAS CHANGED ONCE

Variables that do not come from the questionnaire are given descriptive names. These are the establishment identifier (*betnr*), the wave identifier (*welle*), the survey date (*kalenderwoche*, *jahr*), the weighting factors (*hrf_kal*, *hrf_kal_refined*), a stratification indicator (*strata_id*) and an indicator for the economic activity (*WZ_abschn*) according to the section of the official Classification of Economic Activities (Klassifikation der Wirtschaftszweige) 2008 (Statistisches Bundesamt 2009). Further variables included are the stratification variables (*grkl_admin*, *WZ_ziehung*) and the location (*region*) differentiated according to eastern and western Germany based on the information contained in the establishment file. The class variables *grkl_fb_aktuell* and *grkl_fb_feb2020* were generated from the information contained in *pe01100* and *pe02000*.

Table 7: Variables carried forward

Variable name	Variable label	Notes
betriebseinheit01100	Correct business unit contacted	
standort01100	Only business site or one of several?	
standort02100	Head office or subsidiary / branch office?	
pe01100	Number of employees in February 2020 according to respondent	
pe04101	Establishment currently training apprentices	Instead of carrying the variable forward in pe04101, for all respondents the information in pe02000 (number of employees), pe04100 (currently training apprentices yes/no) and pe05000 (number of apprentices at present) was used to generate the indicator variable “training establishment”, which reflects the training activity status at the time of the respective survey wave.
pe09100	Establishment authorised to train apprentices	Available only from wave 8 onwards and not for all panel cases.
struk02100	Qualification structure: employees for menial jobs	
struk03100	Number of employees for menial jobs	
struk04101	Exporting establishment	Consistently surveyed and accordingly carried forward only from wave 2 onwards.
struk05100	Foreign ownership	
struk06100	Works council	
wa01100	Direction of economic effects up to end of May 2020	
wa02100	Strength of negative economic effects up to end of May 2020	
wa06100	Strength of positive economic effects up to end of May 2020	

4.2 Data validation and correction

Part of the data validation procedure is performed automatically via the software used by the survey institute during the telephone interviews. This ensures that all the details provided fall only within the previously defined range of values. Furthermore, plausibility checks were implemented for some of the responses. For example, establishments that are interviewed for the first time always report the number of employees they had before the crisis (February 2020) as well as the number of employees at the time of the current survey. If there is an overly large discrepancy between the two figures, the establishment is asked a verification question to either confirm the validity or correct the information.

During the data collection process, different types of item non-response were labelled with -7/9997 “does not apply”, -8/-9998 “don’t know”, -9/-9999 “no response” and other numerical codes. In the course of the data preparation process, these non-responses were converted into corresponding missing-value categories. An overview is provided in Table 8. Allocating missing codes rather than using numerical codes is intended above all to prevent errors in the evaluation process. In addition, labelling the different types of non-response increases the

analysis potential of the data, making it possible, for example, to include observations with “don’t know” or other non-response reasons separately in analyses.

Table 8: Classification of missing values

Missing code	Term	Notes
.a	Don't know	Is assigned in accordance with the specification in the questionnaire.
.b	No response	Is assigned in accordance with the specification in the questionnaire.
.c	Does not apply	Applies to wa05000/1. Is assigned in accordance with the specification in the questionnaire.
.z	Question not asked	Question was not asked as a result of the filter procedure.
.f	Filtering error, question was not asked	A few establishments were erroneously not asked the question. Concerns wa11001, wa11021, kug15000 (wave 9), struk09000 (wave 10) and wa26100 (wave 14). The error was recognised and remedied early in the fieldwork period.
.h	All employees	Applies only to home07000 and home09000 in wave 4 and home04000, and home15000 in wave 4, 8, 10, 11 and 12. Is assigned if the establishment does not report a definite figure but responds with “all employees”. Could therefore be set to the number of employees at present (pe02000).
.k	Doesn't occur	Applies only to verhalten01000a-verhalten01000h and verhalten02000a – verhalten02000h in wave 6. Is assigned if an establishment reports that the respective situation does not occur in the establishment.
.l	Establishment had to close because of lockdown measures (November until 15.12)	Applies to wa05000/1b in wave 7 and 8, and corrects the filtering according to information in the questionnaire. Is assigned if an establishment responded with “yes” (establishment was closed partially/entirely) in wa11000.
.l	Employee leasing completely unrealistic	Applies to atypisch08000_* in wave 11. Is assigned in accordance with the specification in the questionnaire.
.r	Not relevant for our establishment	Applies to questions in the topic area “ausrek” in wave 7.
.t	Interview after strict lockdown from 16.12.2020 onwards, question therefore not asked	Applies to wa11000, wa14000 and wa15000. Is assigned to all establishments interviewed later than 16.12.2020 in wave 7, as the corresponding question was no longer asked from that time onwards due to its content.
.v	Establishment currently does not train apprentices	Applies to pe05001 in wave 6 and ausb03001 in wave 7. Is assigned in accordance with the specification in the questionnaire.
.g	There are no contractual working hours	Applies to struk07000, struk09000, struk10000, struk12000 in wave 10. Is

Missing code	Term	Notes
		assigned in accordance with the specification in the questionnaire.
.t	Interview before adding the question, therefore not asked	Applies to kug36000, kug37000, kug38000 in wave 12. Assigned to all establishments surveyed before 21 April, as the question was only asked after that date.
.x	Not possible	Applies to org06000* in wave 13. Is assigned in accordance with the specification in the questionnaire.
.p	Already existed before the pandemic	Applies to psych02000* in wave14. Is assigned in accordance with the specification in the questionnaire.

4.3 Weighting

The panel contains two provisional weighting factors, *hrf_kal* and *hrf_kal_refined*. *Hrf_kal* was developed as a simple procedure at the start of the project and was used until March 2021 for several publications. As a result of a considerably higher willingness to participate among establishments that had submitted a claim for short-time work allowance to the Federal Employment Agency, the weighting factor *hrf_kal_refined* was developed to eliminate this effect and to remedy other shortcomings of *hrf_kal*. The weighting factors yield largely similar results, with the exception of analyses of short-time work.

For *hrf_kal*, the probability of an establishment being included in the sample was first calculated for each separate stratum, measured in terms of the number of contacted establishments as a proportion of all the available establishments. The stratification variables (see also Table 6) and the region (east/west) served as the subsequent non-response correction. Finally, a calibration was performed using the same variables.

In the case of *hrf_kal_refined*, the various selection steps (sampling, telephone numbers that cannot be found, exclusion of establishments, duplicate telephone numbers, no contact due to short fieldwork period, establishment that could not be contacted and the willingness of establishments to cooperate) were modelled step-by-step with a series of logistic regressions, whereby an improved correction of non-response processes was the main goal. The covariables used in addition to the stratification variables listed in Table 6 were the economic activity (17 categories), the region (east/west), the number of employees subject to social security, the age of the establishment in years, the existence of a reporting gap in the administrative data in 2019, the time when the (refreshment) sample was drawn, the size of the municipality in which the establishment is located and – most importantly – the number of months in which an establishment claimed short-time work allowance between April and August 2020. As before, the subsequent calibration was performed using the stratification variables and the region.

The weighting factor *hrf_kal_refined* takes into account the fact that establishments that claimed short-time work allowance in spring/summer 2020 are more likely to be willing to participate in the survey. This insight was a decisive reason for creating this improved weighting factor. However, there is reason to believe that although the correlation between an establishment having claimed short-time work allowance and being willing to participate

was particularly high in the first waves (August/September 2020), it diminished in later waves. It can therefore be assumed that the non-response correction is more successful in the first waves than in waves that were conducted several months later. This may lead to poor comparability of the results over time when comparing waves that are far apart in time. This limitation does not apply for *hrf_kal*.

As a result of the explanations provided above, *hrf_kal* and *hrf_kal_refined* serve only as provisional weighting factors. Final weighting factors are envisaged for the end of the survey.

4.4 Linkage with administrative establishment data (BHP)

For establishments that consented to a linkage with administrative data held by the Federal Employment Agency, additional variables from the Establishment History Panel (Betriebs-Historik-Panel - BHP) are available. These include, for example, general information on the establishment and workforce structure. As of the second quarter of 2021, cross-sectional data from the BHP are available until 2019 (reference date: 30 June). Detailed information and documents concerning the content and the development process of these data can be found on the website of the Research Data Centre: (https://fdz.iab.de/de/FDZ_Establishment_Data/Establishment_History_Panel.aspx).

Once access to the panel “Establishments in the Covid-19 Crisis” has been granted, the BHP data of all the establishments that consented to the linkage are included by default and are available in a separate file. The two data sources can be merged into one dataset via the key variable “*betnr*” (see data example 1). In this process the following must be taken into account:

- For 0.2 percent of the establishments surveyed a linkage is of no use as these establishments did not provide information about their own establishments but about a superordinate unit. The data gathered in the survey therefore do not refer to the same study unit. The variable “*betriebseinheit01100*” serves as an indicator for these establishments.
- Establishments that were newly founded after the reference date of the BHP cannot be linked.
- Establishments whose establishment number changed between the reference date of the BHP and the survey period of the study³ cannot be linked either.

A total of 92.85 percent of the establishments gave their consent for a linkage with administrative data. The consent for data linkage was always asked for at the beginning of the first interview and, if necessary, again at the end of the interview. Establishments that initially refused to agree to data linkage were asked for their consent again if they participated more than once. Establishments that gave their consent for data linkage were able to withdraw their consent at any time. In the datasets of all the individual waves the establishment’s final status regarding consent for linkage is always noted and valid.

³ A detailed description of the establishment number logic can be found in Fritsch/Brixy 2004 and on the website of the Federal Employment Agency (https://www.arbeitsagentur.de/datei/dok_ba015200.pdf)

The BHP extension file is made available on application if valid reasons are stated. The variables contained in the BHP can be found on http://doku.iab.de/fdz/access/BHP_Variablen.pdf.

Data example 1: Linkage with administrative establishment data

```
use ${orig}\BeCovid_W1_v2.dta, clear

merge 1:1 betnr using ${orig}\BeCovid_W01-W14_v1_bhp_7519_v2_2019.dta, keep(1 3)
```

4.5 Anonymisation

For reasons associated with data protection legislation, both the data of the panel on “Establishments in the Covid-19 Crisis” and those of the BHP are weakly anonymised by default. Particularly sensitive variables that may render the surveyed establishment identifiable are available only in coarsened form. This applies to information regarding the place of work (federal state) and the economic activity (3-digit code). If necessary, the original information can also be made available. For this, however, the analysis objective must be stated explicitly in the application for data access. The variables that are regarded as especially sensitive in terms of data privacy are:

Administrative establishment data:

- Place of work – district (Kreis) (ao_kreis)
- Economic activity – sub-class of economic activity (5-digit code) (w08_5)
- First appearance of establishment number (grd_dat)
- Last appearance of establishment number (lzt_dat)

5 Information for data use

5.1 Preparations for data analysis

For a representative analysis of the data it is recommended to use the “*hrf_kal_refined*” weighting factor. In Stata the data can be prepared using the following line of code:

Data example 2: Preparation for data analysis

```
svyset betnr [pw= hrf_kal_refined], strata(stratum_id)
```

All the individual waves can be merged to form a panel dataset. The variable “betnr” serves as an establishment’s consistent identifier across the individual waves. It must be taken into account that only approximately 60 percent of the establishments in each wave are panel respondents (see Table 4). Moreover, there is no guarantee that an establishment will continue to participate in all subsequent waves.

5.2 Stata evaluation example

Due to the high-frequency panel design, the survey data are particularly suitable for analysing time series. In data example 3 a time series is generated for the proportion of establishments that hire staff, and the procedure is demonstrated on the basis of the software Stata.

Data example 3: Time series of the proportion of establishments that have hired staff

```
//Generation of a time series for establishments' recruitment activity

*Install ado estout for estpost command (only necessary once)

fdzinstall estout

*Local for maximum wave

local weakt = 14

*Estimate for each wave the proportion of establishments that hired staff in the past
three weeks (variable only available from wave 2 onwards).

forvalues x = 2(1)`weakt' {

*Load dataset

if(`x' <= 9){

use "$orig\BeCovid_W`x'_v2.dta", clear

}

}
```

```

else{

use "$orig\BeCovid_W`x'_v1.dta", clear

}

*Set-up of weighting

svyset betnr [pw = hrf_kal_refined], strata(stratum_id)

*Estimate

di "Unweighted case numbers in wave 'x'"

tab hire07000, matcell(nobsW`x') //Unweighted case numbers

di "Weighted case numbers in wave 'x'"

estpost svy: tab hire07000, ci percent //Weighted case numbers

*Initialising outcome matrix

mat W`x' = `x' //Wave

mat eeW`x' = e(b) //Proportion

mat elbW`x' = e(lb) //Lower bound of 95%-CI (see set level)

mat eubW`x' = e(ub) //Upper bound of 95%-CI (see set level)

*Combine individual vectors

mat resW`x' = W`x'', eeW`x'[1,1]', elbW`x'[1,1]', eubW`x'[1,1]', nobsW`x'[1,1]'

*Designate column names

mat colnames resW`x' = wave share lowerbound upperbound nobs

}

*Generate outcome matrix for all waves for data protection check

local mn = "resW2"

forv x = 3(1)`weakt' {

local mn = "`mn'" + "\resW`x'"

}

mat resMat = `mn'

di "Outcome matrix (nobs=unweighted number of observations):"

```

```

matrix list resMat

*Generate variables from outcome matrix

svmat resMat, names(col)

*Generate line-plot (proportion of establishments that have hired staff in the past
three weeks per wave)

gr tw (line share wave, lp(dash)) (rcap lowerbound upperbound wave), ylab(0(10)30)
legend(order(1 "hired in the last 3 weeks")) graphregion(fcolor(white) lcolor(white))
xtitle("") ytitle("proportion of the establishments (weighted) as %") xlabel(2 "25.08
- 04.09" 3 "14.9 - 25.09" 4 "05.10 - 16.10" 5 "26.10 - 09.11" 6 "16.11 - 27.11" 7
"07.12 - 18.12" 8 "18.01 - 29.01" 9 "08.02 - 19.02." 10 "01.03 - 15.03" 11 "22.03 -
08.04" 12 "19.04 - 30.04" 13 "10.05 - 27.05" 14 "07.06 - 18.06",angle(40)) title("Hiring
during the Covid-19 crisis") subtitle("Establishments in Germany (2020/21)")

graph export ${log}/Einstellungen_w2_bis_w14.png, replace

```

5.3 Data access and additional material

The data product made available to external researchers includes all the individual waves of the panel “Establishments in the Covid-19 Crisis” conducted to date, as well as establishment data from the Establishment History Panel for 2019.

The weakly anonymous data can only be evaluated in the context of research visits to the Research Data Centre (FDZ) with subsequent remote data access.

In order to use the data it is necessary first to submit a request to the FDZ. Details regarding requests to use the dataset and the data processing possibilities can be found on the FDZ website.

References

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

Ganzer, Andreas; Schmidlein, Lisa; Stegmaier, Jens; Wolter, Stefanie (2020): Betriebs-Historik-Panel 1975-2019. (FDZ-Datenreport, 16/2020 (de)), Nürnberg, 113 S.

Statistisches Bundesamt (Hg) (2009): Klassifikation der Wirtschaftszweige. 2008, Wiesbaden.
URL: <https://www.destatis.de/DE/Methoden/Klassifikationen/Gueter-Wirtschaftsklassifikationen/klassifikation-wz-2008.html> (Zugriff: 06.04.2021)

Imprint

FDZ-Datenreport 13|2021

Date of publication

27 September 2021

Publisher

Research Data Centre (FDZ)
of the Federal Employment Agency (BA)
at the Institute for Employment Research (IAB)
Regensburger Straße 104
90478 Nuremberg

All rights reserved

Reproduction and distribution in any form, also in parts, requires the permission of FDZ.

This document can be downloaded from

http://doku.iab.de/fdz/reporte/2021/DR_13-21_EN.pdf

Documentation version

BeCovid_W01-W14_DE_v1_dok1, DOI: 10.5164/IAB.FDZD.2113.en.v1

Dataset version

BeCovid W01-W14 v1, DOI: 10.5164/IAB.BeCovid_W01-W14.de.en.v1

All publications in the series “FDZ-Datenreport“ can be downloaded from

https://fdz.iab.de/de/FDZ_Publications/FDZ_Publication_Series/FDZ-Datenreporte.aspx

Website

<https://fdz.iab.de/>

Corresponding author

Patrick Gleiser
Tel.: (911) 179 7987
Email: patrick.gleiser@iab.de