Kolja Briedis | Almuth Lietz | Kai Mühleck | Uwe Ruß | Percy Scheller | Isabelle Schiffer-Fiedler | Ulrike Schwabe | Moritz Seifert | Carola Teichmann | Susanne de Vogel | Robert Birkelbach | Ute Hoffstätter | Tina Oestreich | Henrike Steins | Anne Weber

Nacaps 2018

Data and methods report for data package version 3.0.0 of the National Academics Panel Study 2018 (1st-6th wave)

nacaps

fdz.DZHW.
Research Data Centre for
Higher Education Research and Science Studies

Table of Contents

List of tables/figures

1 Content and design of the study

Nacaps 2018 is part of the survey series *National Academics Panel Study (Nacaps)*, a longitudinal study that, for the first time, systematically collects data on the career and life courses of highly qualified individuals in Germany using standardised online surveys addressed to doctoral candidates and doctoral holders – for research, higher education institutions and science policy (Wegner & Briedis, 2020).[1] Nacaps is conducted by the German Centre for Higher Education Research and Science Studies (DZHW) and was funded by the Federal Ministry of Education and Research (BMBF) until 2024.

The Nacaps study series has a multi-cohort panel design, in which cohorts of doctoral candidates, i.e. persons who were registered for a PhD/doctorate at a German university on a specific cut-off date, are contacted every two years and subsequently surveyed annually (cf. Figure 1). So far, there are cohorts of doctoral candidates with a cut-off date in December 2018, 2020 and 2022, which all were surveyed for the first time in the following year. To refresh and partially supplement the data, doctoral holders who finished their PhD/doctorate in 2023 were also surveyed once in 2024. Furthermore, there are also plans for a new cohort of doctoral candidates, which will be surveyed for the first time in 2027. In 2028, doctoral holders who finished their PhD/doctorate in 2027 are supposed to be integrated into the panel as well.

Figure 1: Nacaps survey design

---
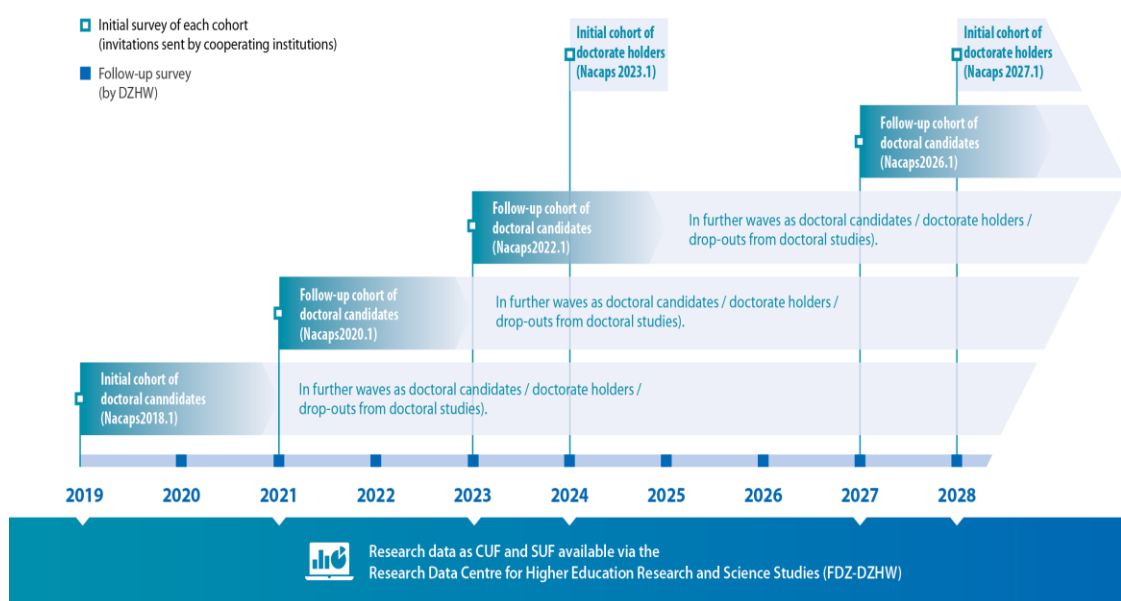
[1] Current information on Nacaps can be found on the project website (https://www.nacaps.de).

Nacaps 2018 is the first cohort of doctoral candidates and marks the start of the study series. The population includes all those who were registered for a PhD/doctorate at a German university on the cut-off date of 1 December 2018. The first survey wave took place from February to May 2019. The follow-up survey waves 2-6 took place at annual intervals in spring.

This means that Nacaps 2018 is characterised by the fact that some of the data collection took place during the Covid19 pandemic. Figure 2 (adapted from Rußmann, Netz, & Schwabe, 2025) shows the annual surveys of Nacaps 2018 (upper part; waves 1-5 only) and the course of the Covid19 pandemic (lower part). The initial survey (wave 1) took place in spring 2019 <u>before</u> the Covid19 pandemic. The two subsequent surveys (waves 2 and 3) took place in spring 2020 and 2021 <u>during</u> the Covid19 pandemic, with the second wave coinciding almost entirely with the first lockdown. From the fourth wave onwards, data collection took place <u>after</u> the Covid19 pandemic. The data from Nacaps 2018 is therefore suitable for analysing the effects of the Covid19 pandemic using a before-and-after design.

Figure 2:   Annual surveys of Nacaps 2018 and the course of the Covid19 pandemic
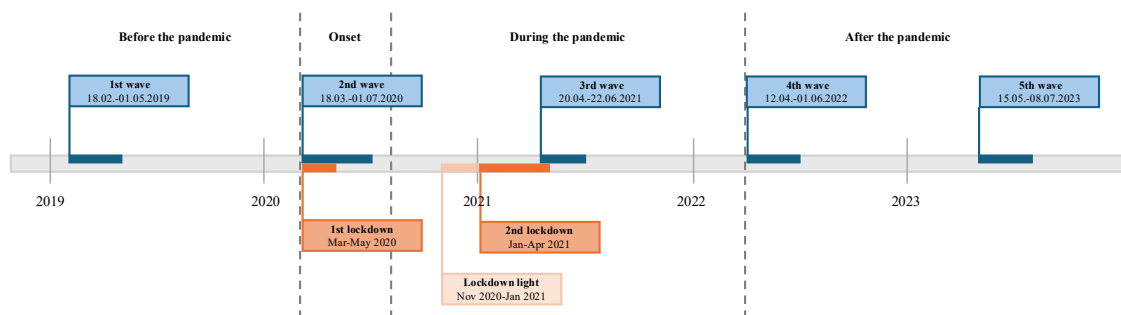
Source: Rußmann, Netz, & Schwabe (2025), p. 650

In addition to basic data for educational reporting, the data contains detailed information on the learning and development conditions during the PhD/doctorate phase as well as on the career and life courses of doctoral holders and drop-outs from doctoral studies following their PhD/doctorate phase. In addition, the data contains a range of personality traits (including Big Five, self-efficacy, internal/external locus of control), information on the health situation and socio-biographical and educational background information. The survey programme is supplemented by wave-specific focus topics (e.g. science communication, Covid19 pandemic). This results in a large, previously unavailable data source for analysis for higher education research and science studies.

2   Survey instrument

In all survey waves of Nacaps 2018, the survey instrument was a standardised online questionnaire in German and English. The questionnaire was programmed and the survey was conducted using the DZHW online survey software *Zofar*.

Chapter 2.1 describes the cognitive and quantitative pretests carried out before the initial survey to test and optimize the survey instrument. Chapter 2.2 presents the central contents of the survey instruments.

2.1   Pretests

[Aim and procedure] The survey instrument was first tested using a cognitive and then a quantitative pretest before the initial survey. The aim of the cognitive pretest was to validate whether the survey instrument was suitable for the target group of doctoral candidates, whether newly developed items were equally suitable for different groups of doctoral candidates and whether they were comprehensible. The aim of the quantitative pretest was to test the survey instrument in the field under similar conditions to the main survey. On the one hand, the validity and reliability of newly developed instruments were to be tested. On the other hand, the duration of the survey, the structure of the questionnaire, the general feasibility of the online survey and the effectiveness of the incentivisation were to be tested.

[Cognitive pretest] From the 6th to the 15th of December 2017, 17 cognitive interviews with doctoral candidates were conducted face-to-face at the Hanover and Berlin locations of the DZHW by four project staff members of Nacaps. Each interview was scheduled to last 90 minutes, with the actual length of the interview ranging from 72 minutes to 110 minutes. The interviewees received 50 euros as a remuneration for their time and effort. Depending on the specific questions to be tested and the respective interest in the insights, interview techniques such as *probing* (e.g. information retrieval probing, comprehension probing), *confidence rating* and *general probing* were used (Lenzner, Neuert, & Otto, 2015; Prüfer & Rexroth, 2000; Weichbold, 2019). People from different subject groups and both women and men were interviewed (Table 1). Various questions and item batteries were tested on the topics of *path to the PhD/doctorate topic*, *types of PhD/doctorate*, *social networks and support*, *scientific activity* and *supervisor relationship*. The sequence of the topics was varied in two different versions.

Table 1:   Gender and research area (DFG) of the interviewed doctoral candidates

Research area	Gender		Total
	female	**male**	
Humanities and Social Sciences	4	1	5
Life Sciences	3	0	3
Natural Sciences	3	2	5
Engineering Sciences	0	4	4
Total	10	7	17

[Quantitative pretest] For the quantitative pretest, 10,089 names and email addresses of doctoral candidates and scientific staff without a PhD/doctorate were gathered on freely accessible websites of 29 German higher education institutions entitled to award PhDs/doctorates from January to March 2018. The questionnaire was programmed and the survey was conducted using the DZHW online survey software Zofar. The field phase began on the 9th of May 2018 and ended on the 31st of May 2018, during which two reminders were sent out (on the 17th and 23rd of May 2018). Of the 10,089 people contacted, 1,692 took part in the survey to completion and 3,109 filled out the survey partially. The response rate was therefore 16.7 and 30.8 percent respectively.

2.2   Contents of the survey instruments

The contents of the survey instruments are first described for the initial survey (cf. chapter 2.2.1) and then for the five follow-up surveys (cf. chapter 2.2.2).

It should be noted that some questions from the survey instruments were used secondarily from other instruments – some unmodified, some modified (cf. appendix 10.1).

2.2.1   Initial survey

At the beginning of the questionnaire in the first survey wave, respondents were asked about their current doctoral status[2]. Depending on their doctoral status, different questions were displayed in the further course of the survey.

The questionnaire was divided into six thematic blocks. *Information about the PhD/doctorate (A)*, *PhD/doctorate thesis and PhD/doctorate conditions (B)*, *everyday work in the PhD/doctoral phase (C)*, *personal life situation and career goals (D)*, *sociodemography and previous stages of education (E)* and finally aspects such as *panel willingness, newsletter, raffle and contact data (F)* were recorded.

[Information about the PhD/doctorate] The thematic block on *information about the PhD/doctorate* comprised questions on motivation at the start of the PhD/doctorate, the start date in terms of content and the date of registration at the higher education institution, the name and location of the higher education institution, reasons for the choice of the higher education institution, cooperation with higher education institutions abroad or universities of applied sciences, and on the doctoral subject. If the PhD/doctorate had already been completed, the date of completion was also asked; if the PhD/doctorate was quit, the date and reasons for discontinuation were recorded.

[PhD/doctorate thesis and PhD/doctorate conditions] The thematic block on *PhD/doctorate thesis and conditions* contained questions on cooperative forms of PhD/doctorates in research projects, on the PhD/doctorate topic, on subjectively perceived PhD/doctorate progress, on structured doctoral programmes, on scholarships, on courses and lectures (including satisfaction with them), on cooperation with external organisations, on the form of the PhD/doctorate (monograph vs. cumulative doctorate), on intentions to drop out, on the existence of a PhD/doctorate or supervision agreement and on the number of supervisors. It also included questions about the doctoral supervisor (affiliation, gender, formal supervision status, position, frequency of exchange, supervision style, satisfaction and likeability rating), good scientific practice, the source of funding and specific questions for doctoral candidates working at a higher education institution or research institution.

[Everyday work in the PhD/doctoral phase] The thematic block on *everyday work in the PhD/doctorates phase* consisted of questions on employment (such as fixed-term contracts, scope of employment), qualification objectives and the thematic relation to the doctoral project. Furthermore,

---

[2] With the four possible answers: "I am doing a PhD/doctorate", "I have completed the PhD/doctorate", "I have interrupted my PhD/doctoral project" and "I have quit my PhD/doctoral project".

questions were asked about specific academic activities, daily motivation to work on the doctorate, the most recent cooperative projects with other researchers or partners outside academia, the number of contact hours and stays abroad during PhD/doctorate.

[Personal life situation and career goals] The thematic block on *personal life situation and career goals* covered general life satisfaction, included questions on partners and parenthood, satisfaction with work-life balance, general state of health, physical and mental well-being and the presence of a disability. The Big Five (five factor model), general willingness to take risks, general locus of control and general self-efficacy expectations were also recorded as psychological constructs. Furthermore, instruments were developed and used to measure the importance of certain job characteristics, the attractiveness of various positions, the intention to remain in the scientific sector, sector preference, the striving for a professorship, the desire for permanent employment in academia and the specific self-efficacy expectations regarding a job in or outside academia.

[Sociodemography and previous stages of education] The thematic block *sociodemography and previous stages of education* contained questions on the respondents' gender, year and place of birth and nationality/nationalities on one hand as well as the place of birth and nationality/nationalities, school-leaving certificate, vocational training, professional status, and educational aspirations of their parents on the other hand. In addition, the place and grade of university entrance qualification and information on the studies that entitled to a PhD/doctorate were collected.

[Panel willingness, newsletter, raffle and contact data] The survey concluded with questions on panel willingness, for contacting to inform about the Nacaps results and for participation in the raffle as well as a request for contact details and the opportunity to provide comments.

2.2.2 Follow-up surveys

The follow-up surveys covered certain core topics in each wave, although the specific questions relating to each core topic varied in some cases. Also, information from previous waves was included in the current survey ('preload') in order to be able to build on it. Additionally, specific topics only appeared in individual waves. The exact content of the core and wave-specific topics is explained in more detail below.

Core topics

[General information on the PhD/doctorate] In all five survey waves, respondents were asked about their current PhD/doctorate status (doing a PhD/doctorate, interrupted, thesis or all relevant papers submitted for examination (from the third wave onwards), completed, quit) at the beginning of the survey. If the PhD/doctorate was completed following the previous survey, respondents were also asked about the date of the completion, changes to the dissertation topic, the form of dissertation (monograph vs. cumulative doctorate) and its publication as well as the final doctoral grade. If the PhD/doctorate was quit/interrupted, the reasons for the discontinuation/interruption were recorded in the second wave and the date of discontinuation was recorded in all five waves, as well as whether the higher education institution was informed about the discontinuation/interruption. For respondents who completed or discontinued their PhD/doctorate following the previous wave, the doctoral university and the (intended) doctoral subject were also recorded (again). From the third wave onwards, respondents who had completed their PhD/doctorate since the last survey were asked about the date on which they received their doctoral certificate and whether they had received an award for their doctoral thesis. Information was also collected on the reviewers (number, institution, formal status of reviewers at universities of applied sciences, supervision status) and the self-assessed reputation of the main reviewer and the institution.

[PhD/doctorate topic/conditions] All five survey waves contained questions on the subjectively perceived progress of the doctorate, expectations of success and intentions to drop out (including reasons), the number of supervisors, the main supervisor (change, affiliation, gender, formal supervisor status, position, frequency of exchange, supervision style, satisfaction and likeability rating) and course attendance (in wave 2 including satisfaction). The second wave also included an instrument consisting of several questions to measure the learning environment in the doctoral phase as well as questions on support regarding good scientific practice and non-financial sponsorship through support programmes. In addition, in the third and fourth waves, questions were asked about cooperation with partners within and outside academia (in wave 3 including cooperation with companies and external organisations) and about teaching commitments. From the fourth wave onwards, there were also questions about the membership in a doctoral programme; only in the fourth wave there were more detailed questions about the type of doctoral programme and details of membership.

[Scientific activities during the PhD/doctorate] In the second wave, questions were asked about the publication media used during the PhD/doctorate phase. From the third wave onwards, the scientific activities during the PhD/doctorate were recorded in more detail with questions on the number and format of scientific publications and contributions in consumer and online media, number of conference visits and contributions, reviews as well as applications for third-party funding and patents (submitted/approved).

[Mobility experiences] From the third wave onwards, information on stays abroad for PhD/doctorate related reasons or other academic purposes since the beginning of the PhD/doctorate were recorded using a table (type of stay, country, date of beginning, duration). The mobility intentions/plans (type of stay, country) were also surveyed. In waves 3 and 4, these questions were only asked to respondents who stated in the respective wave that they had completed or quit their PhD/doctorate. Thus, in these two waves, stays abroad since the start of doctorate were recorded once for these status groups. From wave 5 onwards, <u>all</u> respondents were asked about their stays abroad and plans to go abroad, with the exception of those who had already quit their PhD/doctorate in the previous waves. Originally, respondents with an ongoing, interrupted or already in previous waves completed PhD/doctorate were supposed to receive a different question text and indicate their stays abroad <u>since the last survey</u>. Unfortunately, a transmission error occurred, so that all respondents received the same question text and were asked (in some cases again) about their stays abroad <u>since the start of doctorate</u>, regardless of whether they had already completed their PhD/doctorate in the previous wave or had an ongoing or interrupted PhD/doctorate. This was the case in both wave 5 and wave 6.

[Financing and employment] In all five survey waves, questions on financing of living expenses of doctoral candidates and interrupters were asked first. These included sources of funding, monthly net income and self-assessed financial security. Additional questions were asked of scholarship holders, those employed at universities and recipients of parental allowance ("Elterngeld"). For doctoral holders and drop-outs the employment status was surveyed first. Those who were not employed were then asked about their financial sources and net income. For all employed persons (doctoral candidates, interrupters, doctoral holders and drop-outs), more detailed questions were asked about the period of employment, fixed-term contracts, scope of employment, the number of contracts, gross monthly income, company size, sector, adequacy, job position, academic personnel category, management responsibility and job satisfaction, as well as the place of work from the third survey onwards. In addition, several questions were asked about the scientific relation of the employment. In addition, information was collected from doctoral candidates on the PhD/doctorate-relation of their employment (e.g. PhD/doctorate as a contractually defined qualification objective, proportion of working time, thematic relation).

[Personal life situation] The five survey waves each contained various questions on partnership and parenthood (in waves 2 and 3 including a question on difficulties regarding family planning), general health and physical and mental well-being. From the third wave onwards, the place of residence was also surveyed.

[Attitudes, self-assessment and goals] In all five survey waves, general life satisfaction, satisfaction with work-life balance, general self-efficacy expectations and the TEN/FLEX instrument were recorded (cf. Brandtstädter & Renner, 1990). In addition, instruments were developed and used to record the importance of certain job characteristics (only in wave 2), the attractiveness of various jobs, the intention to remain in the science sector, the sector preference, the striving for a professorship, the motivation for a doctorate (only in waves 3 and 4), the desire for permanent employment in academia, plans for self-employment and the specific self-efficacy expectations with regard to an activity within or outside academia.

[Conflicts during the doctoral phase] From wave 6 onwards, a set of questions was added on conflicts experienced during the doctoral phase. These questions concerned conflicts experienced personally, the people involved, discrimination based on characteristics (ethnicity, religion, etc.) and experiences of sexual harassment and bullying.

[Contact data] All five survey waves concluded with questions to request or update contact details and the opportunity to provide comments.

Wave-specific topics

[Political participation] In the second survey wave, six questions focused on the topic of political participation. Respondents were asked about their interest in politics, their self-assessed aptitude for political commitment, their perceived opportunity to have an influence on politics and to take part in politics, their assessment of the extent to which politicians take their concerns into account and, finally, their own political activities in the last twelve months.

[Further education] In the second survey wave, two questions were also used to determine whether the respondents had completed a practical training phase or a professional training organised by a professional body.

[Grandparents] In the second survey wave, four questions were also asked about the grandparents' educational background, the relationship and the frequency of contact with the grandparents.

[Panel Commitment] The third survey wave was preceded by a question to record the panel commitment of the participants.

[Science communication] In the third survey wave, another focus was placed on the topic of science communication with five questions. In addition to the question of whether the survey participants engage in science communication, the type of activities, the communication media and the reasons for participation or non-participation in science communication were also surveyed.

[Covid19 pandemic] In the fourth survey wave, the focus was on working conditions in the context of the Covid19 pandemic. The survey asked about the scope of work, the place of work, the impact of the Covid19 pandemic on working conditions, the assessment of how the pandemic was handled by the employer and how the respondents were personally affected by the pandemic.

3 Population and sampling procedure

[Population] Based on the definition in the federal government's reporting, young researchers are defined as university graduates who, after completing their studies, take part in scientific activities or do a PhD/doctorate inside or outside of higher education institutions or non-university research institutions, as well as doctoral holders who are pursuing or intend to pursue a scientific activity inside and outside of higher education institutions and non-university research institutions. The exact definition of who is considered a doctoral candidate had not been standardised for a long time. Nacaps is therefore based on the definition of the Higher Education Statistics Act[3] and the doctoral candidates statistics introduced by the Federal Statistical Office in 2017. Accordingly, the population of doctoral candidates in Nacaps 2018 consisted of those people who were registered for a PhD/doctorate at a German university on the cut-off date of 1st December 2018. According to the Federal Statistical Office, the number of doctoral candidates in Germany defined as such totaled 173,339 in 2018 (Federal Statistical Office, 2019c).

[Total universe/Complete enumeration] In the first step, all 155 universities entitled to award PhDs/doctorates (as of October 2017) were invited to participate, followed by a second step in which all (registered) doctoral candidates from the 53 participating higher education institutions were invited to take part in the survey.[4] As neither higher education institutions entitled to award PhDs/doctorates nor doctoral candidates were selected or drawn, the chosen procedure is a complete enumeration. Non-participating higher education institutions and doctoral candidates are treated as nonresponse.

---

[3] Doctoral candidates are persons who have received written confirmation from an institution entitled to award doctorates that they have been accepted as a doctoral candidate at this institution (Sect. 5 (1) HStatG)

[4] It is possible that persons were still registered at one of the partner higher education institutions at the time they were contacted but had in fact already transferred to another higher education institution and the information in the survey already refers to this new institution.

4 Implementation of the survey

4.1 Initial survey

[Establishing contact and maintaining addresses] Of the total of 53 participating higher education institutions, 48 participating higher education institutions sent out invitations and reminders themselves (address mediation ("Adressmittlung")) and three higher education institutions sent contact details to the DZHW, which took over the invitation and reminder dispatch (address transmission). Two participating higher education institutions were neither able to send contact details to the DZHW nor send an invitation to the survey directly. For this reason, the doctoral candidates at these two higher education institutions were asked to register for the survey on a separate website. The DZHW then invited the registered doctoral candidates to take part in the survey.

At the participating higher education institutions, invitations and reminders were mostly sent out by central offices such as graduate centres and evaluation offices, which in many cases were already known to the doctoral candidates at the higher education institution. To ensure that invitations and reminders were sent out as uniformly as possible, the DZHW provided various materials to those higher education institutions that invited their doctoral candidates to take part in the survey themselves. These included a handout and a video tutorial on sending invitations and reminders as well as mail merge letter templates for the pre-announcement, invitation and reminder letters. In addition, the participating higher education institutions were invited to a workshop on conducting the survey at the DZHW in Hanover.

[Survey documents] The survey documents included a letter of pre-announcement, a letter of invitation and three reminders. Survey participants received a personalised link to the online questionnaire with the invitation email and all subsequent reminders. The survey was programmed and conducted using the DZHW online survey software *Zofar*. Reminders were only sent to those people who had not yet fully participated in the survey. Information on data protection was provided at the beginning of the online questionnaire when asking for consent to the data protection regulations. All letters were made available to the participating higher education institutions by the DZHW and were only slightly modified by inserting the customary form of address and the respective greeting phrasing of the sending higher education institution.

[Field phase] The field phase took place from the 18th of February to the 1st of May 2019. Pre-announcement, invitation and reminders to participate in the survey were sent out on the following days (cf. Table 2).

Table 2:	Dispatch dates of the survey documents of the initial survey

Date	Survey documents
02/11/2019	Pre-announcement
02/18/2019	Invitation / start of the field phase
03/05/2019	1st Reminder
03/20/2019	2nd Reminder
04/04/2019	3rd Reminder

At some higher education institutions, the invitation was sent out a few days later than planned. At five higher education institutions, the invitation email could only be sent on the 5th of March 2019 due to organisational reasons; in another case, the invitation email was only sent on the 22nd of March 2019. Due to these delayed invitations, the field phase was not completed until the 1st of May 2019. This ensured a sufficiently long field phase for the doctoral candidates at these higher education institutions and the possibility to send out three reminders as well. A few higher education institutions had spoken out in advance against sending out three reminders and instead only sent out one or two reminders.

[Measures to increase response rate] Several measures were taken to increase the response rate. Firstly, a non-monetary promised incentive in the form of a raffle was offered with the survey invitation and reminders. The following non-cash prizes were raffled off among all participants.

- 1 x Apple iPad Pro 11" with 256 GB worth EUR 1,049
- 2 x Microsoft Surface Go with 64 GB worth EUR 429 each
- 10 x Gift vouchers ("Wunschgutschein") worth EUR 50 each
- 25 x Netflix vouchers worth EUR 25 each

In one case, a participating higher education institution offered its own additional prize draw for its own respondents.

Additionally, the participating higher education institutions were asked to make their doctoral candidates aware of the upcoming Nacaps survey before the field phase began. For this purpose, the higher education institutions were provided with posters, flyers, sample texts for news posts on websites and image material. In addition, the higher education institutions were asked to send the pre-announcement e-mail to their doctoral candidates one week before the start of the field phase. On the one hand, this was intended to make the doctoral candidates aware of the upcoming survey and the project. On the other hand, it was used to test the sending of emails by the higher education institutions. In addition to directly addressing the doctoral candidates, various accompanying public relations measures were carried out to increase the interest of doctoral candidates in the project. In order to do this, a project website (https://www.nacaps.de/) and a Twitter account (@Nacaps_Panel) were set up a few months before the start of the field phase through which contributions relevant to the target group were shared, particularly before and during the field phase.

Furthermore, the invitation and the three reminders were sent out on different days of the week. The invitation was sent out on Monday, followed by the first reminder two weeks later on a Tuesday. The

second reminder a further two weeks later was sent on a Wednesday and the third reminder was also sent two weeks later on a Thursday.

As the implementation of the project on the one hand and the accessibility of the doctoral candidates on the other depend to a large extent on the willingness of higher education institutions entitled to award PhDs/doctorates to cooperate, they were also incentivised with a series of measures. Participating higher education institutions therefore were given access to a data portal that enables higher education institution-specific analyses of their own survey data. Furthermore, the participating higher education institutions were greeted with a welcome package (Nacaps coffee mug, fridge magnet, post-it notes, flyer, poster, etc.).

4.2   Follow-up surveys

[Establishing contact and maintaining addresses] In contrast to the initial survey, the invitation letters for the subsequent surveys could be sent directly to the respondents via the Nacaps email address, as the respondents were asked in the initial survey to consent to being contacted (for further surveys) and to provide their addresses.

During the follow-up surveys, cases of withdrawal, requests to discontinue contact as well as address changes and updates by means of absence notes (in the case of a newly provided address) or non-delivery messages (if possible, recourse to an alternative address) were entered into the address database. In waves 3 and 4, contact campaigns were also carried out before the invitations were sent out, which served, among other things, to maintain addresses.

After the initial survey, most of the respondents also signed up for the Nacaps newsletter, which was sent annually – usually at the end of the year – to all subscribers with information about the project and relevant reviews, publications or conference contributions by project staff.

[Survey documents and field phase] The follow-up surveys, like the initial survey, were conducted online using the DZHW's internal *Zofar* survey software. For all follow-up surveys, the invitation and reminder letters were sent by email with a personalised link to the online questionnaire.

The exact dates of the field phases of the follow-up surveys are shown in Table 3. The second survey wave started in mid-March; from the third survey wave onwards, the start of the field phase was postponed to mid-April for organisational reasons. In wave 5, the field phase started in May.

Dates	2nd wave	3rd wave	4th wave	5th wave	6th wave
Pre-announcement/ contact campaigns	-	Christmas card, by post on 12/11/2020	Pre-announce-ment postcard, by post on 04/04/2022	-	-
Start	03/18/2020	04/20/2021	04/12/2022	05/15/2023	04/17/2024
1st reminder	03/25/2020	04/27/2021	04/20/2022	05/23/2023	04/22/2024
2nd reminder	04/09/2020	05/20/2021	05/05/2022	06/12/2023	05/07/2024
3rd reminder	06/03/2020 ("Restart")	06/16/2021	05/25/2022	06/28/2023	05/22/2024
4th reminder	06/16/2020	-	-	-	-
End	07/01/2020	06/22/2021	06/01/2022	07/08/2023	06/07/2024

Table 3: Dispatch dates of the survey documents of the follow-up surveys

Due to the Covid19 lockdown, the field phase of the second survey wave was extended over four months and a "restart letter" and an additional reminder were sent out to counteract the low response rate in the initial phase of Covid19 pandemic.

Prior to the third survey, a Christmas card was sent by post to all panel participants with a valid postal address. Before the start of the fourth survey wave, a postal pre-announcement was sent to the panel participants with a valid postal address.

[Measures to increase response rate] A constant measure to increase the response rate in the follow-up surveys is incentivisation through an announced raffle (technical prizes and/or vouchers). The prizes were mentioned in every letter. In most cases, a large non-cash prize (such as a fairly manufactured, up-to-date smartphone), two smaller non-cash prizes (such as two audio devices from well-known manufacturers) and ten vouchers with a value of 50 euros each were raffled off.

Furthermore, the additional contacts (Newsletter delivery annually, second wave: field phase extension and additional reminder letter; third wave: Christmas postcard; fourth wave: pre-announcement letter) should motivate more people to participate.

5   Response rate

5.1   Initial survey

The first survey wave of Nacaps 2018 took place at 53 German higher education institutions, with all 155 higher education institutions entitled to award PhDs/doctorates in Germany being asked to take part in the survey. A complete enumeration was conducted at higher education institution level, i.e. all doctoral candidates registered on the cut-off date of 1 December 2018 were invited by the cooperating higher education institutions.

Table 4:	Participation of higher education institutions

Involved	Quantity
Higher education institutions entitled to award PhDs/doctorates	155
Participating higher education institutions	53

A total of 77,509 invitations to the survey were successfully sent out. Respondents of 28,368 questionnaires agreed to the data protection regulations, which results in an (adjusted) response rate of 36.6 percent (Table 5).

Table 5:	Gross and net response rate of the first Nacaps survey wave

Wave 1	
Invitations sent [1]*	80.086
Undeliverable*	2.577
Adjusted invitations sent [2]*	77.509
Questionnaire link opened [3]	29.842
Consent to data protection [4]	28.368
Questionnaire content completed [5]	23.612
Response rate 1 ([3]/[1])	37,3 %
Response rate 2 ([3]/[2])	38,5 %
Response rate 3 ([4]/[1])	35,4 %
Response rate 4 ([4]/[2])	36,6 %
Response rate 5 ([5]/[1])	29,5 %
Response rate 6 ([5]/[2])	30,5%

* It should be noted that the number stated here also includes cases that were additionally invited to the survey for other purposes.

It should be noted that during the course of the panel, some respondents requested that their data be completely deleted, which is why the number of cases in the published data is slightly lower than in the response calculations in Table 5.

Figure 3 shows the development of the response rate over the entire course of the field period as a relative proportion of completed questionnaires. In addition, the times at which the three reminders were sent are marked, which show an overall increase in response.

Figure 3:   Relative proportion of completed questionnaires (in %) over the course of the field period

Approximately 10 weeks in the field
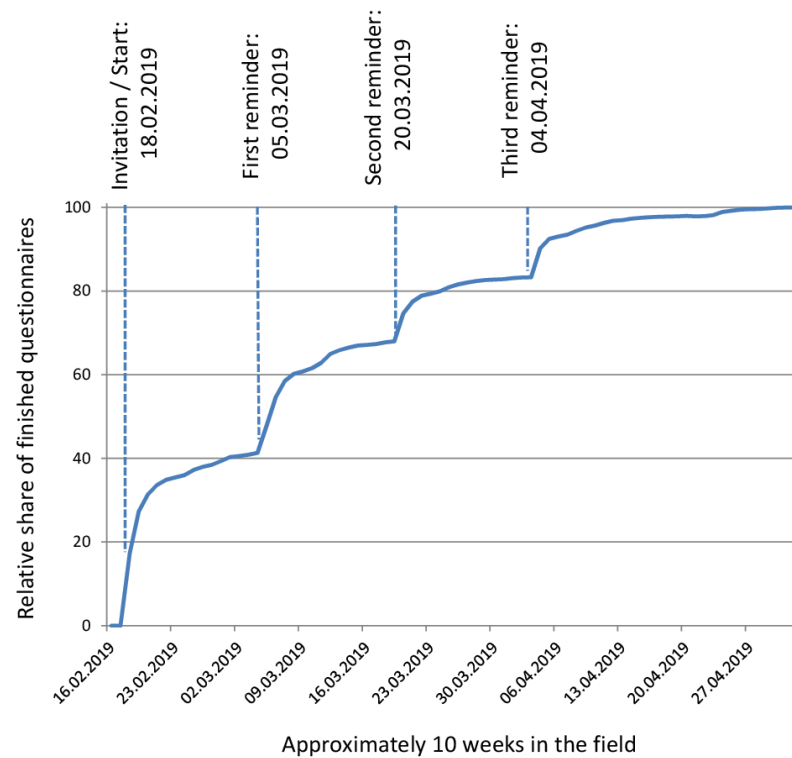
5.2   Follow-up surveys

The response rate of the follow-up surveys was recorded in the table for calculating the response rate of the American Association for Public Opinion Research (AAPOR)[5]. Table 6 shows the main response figures and the response rate according to AAPOR RR1 (minimum rate) and RR2 (including partially completed questionnaires).

In wave 2, invitations were sent to all participants from wave 1 who had given their consent for further surveys in wave 1, provided a valid email address and had not withdrawn their participation in the meantime. There is no general attrition rule for the rest of the panel, i.e. respondents are also invited to the next follow-up survey if they have not participated in any of the waves in the meantime. However, there is still attrition between survey waves, as a few individuals have withdrawn from the study and are therefore no longer contacted.

It should be noted that cases for which a request for complete deletion of data had already been made at the time of the response calculations for Table 6 (cf. chapter 5.1) were no longer included in the calculations. At the same time, it must be taken into account that further data deletions were made after the response calculations, so that there may be differences between the case numbers in Table 6 and the published data.

---

[5] https://aapor.org/response-rates/ (accessed on: 11/20/2025)

Table 6:   Gross and net response rate of the Nacaps follow-up survey waves (according to AAPOR 2016)

Data	2nd wave	3rd wave	4th wave	5th wave	6th wave
Invitations sent [1]	17.247	17.068	16.893	16.806	16.083
Undeliverable	157	271	486	488	210
Withdrawals (during the field phase)	20	24	46	54	49
Survey opened but not filled [2]	473	494	237	311	246
Partially completed (break-off from A01 and before C01) [3]	626	527	387	284	370
Fully completed (break-off from C01 or no break-off) [4]	8.893	6.994	6.915	5.420	4.885
Number of cases ([2]+[3]+[4])	9.992	8.015	7.539	6.015	5.501
of which usable cases ([3]+[4])	9.519	7.521	7.302	5.704	5.255
Response Rate 1 (RR1) is the minimum response rate. ([4]/[1])	51,6 %	41,0 %	40,9 %	32,3 %	30,4 %
Response Rate 2 (RR2) counts partial interviews as respondents. ([3]+[4]/[1])	55,2 %	44,1 %	43,2 %	33,9 %	32,7 %

6   Data preparation

6.1   Data transfer

The survey was conducted using the DZHW's own online survey system *Zofar.* The data from the online survey was delivered by *Zofar* to the Nacaps team as a .csv file and then processed as a Stata file (.dta). The raw data set was stored on a protected server.

6.2   Data checking and cleansing

[Checking the value ranges for numerical open responses during the survey] In the initial survey, value ranges that should not be exceeded or fallen short of were defined within the questionnaire programming for numerical open responses to limit the extent of implausible values from the outset. If the value ranges were exceeded or undercut, the respondents were informed in red that their entry appeared implausible and were asked to correct the value. If no correction was made after the notification, the survey could still be continued. In the subsequent waves, valid value ranges were not defined and only types (numeric or string) were specified; in the third wave, no plausibility checks were carried out at all during the survey.

[Subsequent cleansing of numerical open responses] During data preparation, there was generally no subsequent cleansing of value ranges for numerical open responses. Implausible or impossible entries were retained so that the user can decide how to deal with them[6]. An exception is made for variables for which, according to the completion instructions, only whole numbers or one decimal place should be entered. In these cases, the deviating values were subsequently rounded and the original information was overwritten. In addition, open responses that required the entry of numerical values but contained text information[7] were recoded to numerical values if a clear assignment was possible. Otherwise the missing value "-966 not determinable" (cf. chapter 6.6) was assigned.

[Filter checks] Furthermore, completeness checks were carried out regarding the filter procedure(s) on the one hand and tests for filter violations on the other (test of adherence to filter procedures). As it was possible to change information in the survey by jumping backwards and forwards, it is not possible to recognize in the data which information is correct. Therefore, no changes were made to the information and the decision as to which information should be used for the analyses can be made by the data users themselves.[8] There is an exception in cases where information was initially provided in the questions about health, but the corresponding informed consent was subsequently withdrawn by jumping back. In these cases, the original information was subsequently recoded to "-989 filtered" (cf. chapter 6.6).

[Deletion of cases or information] Some cases or information were removed from the data set during the course of the panel. If respondents only wanted information from individual waves to be deleted, the information for these waves was set to "-995 not participated (panel)" (cf. chapter 6.6), but the

---

[6] It should be noted, however, that the open responses on children (number and years of birth) were checked for plausibility and, if necessary, cleaned prior to aggregation in the course of anonymisation (cf. chapter 8) (without changing the original variables).

[7] As no response types (numeric vs. string) were defined in advance in the survey programming of the third wave, this occurred more frequently in the third wave compared to the other waves.

[8] One way to mark filter violations for data users is to create flag variables. Although filter violations occurred for the majority of the variables, they only ever affected isolated cases. As these additional flag variables would greatly inflate the data sets, the marking of filter violations with flag variables was dispensed with after weighing up the costs and benefits.

cases were generally left in the data set. In the case of a request for complete deletion of the data, this case was removed completely (cf. chapter 5).

6.3 Generation of variables

As part of the data preparation, additional variables were generated (identified by the suffix "_g#" in the variable name, cf. chapter 6.5). These are, firstly, variables with numerical coding of originally open-ended responses[9], secondly, adjusted/cleansed variables (cf. chapter 6.2) and thirdly, variables that were generated for data protection/anonymization reasons (cf. chapter 8). For each generated variable contained in the data, the generation procedure is described in the data set report. In some cases, reference is made to specific lists, which can be found here:

- For higher education institution variables:
  - Waves 1-4: Destatis key list of student and examination statistics, doctoral candidates statistics WS 2018/2019 and SS 2019 (Federal Statistical Office, 2019a): https://www.statistik-nord.de/fileadmin/Dokumente/Erhebungen/Hochschulstatistiken/1_Studen-ten_2_Pr%C3%BCfungen/Schluesselverzeichnis_WS_2018.xlsx, key 2.2 (higher education institutions by type of higher education institution and federal state) and key 2.3 (higher education institutions, numerical) (accessed on 04/16/2025)
  - Wave 5: Destatis key list of student and examination statistics, doctoral candidates statistics and guest student statistics WS 2021/2022 and SS 2022 (Federal Statistical Office, 2022a): https://www.statistik-nord.de/fileadmin/Dokumente/Erhebungen/Hochschulstatis-tiken/1_Studenten_2_Pr%C3%BCfungen/Schluesselverzeichnis_WS_2021.xlsx, key 2.2 (higher education institutions by type of higher education institution and federal state) and key 2.3 (higher education institutions, numerical) (accessed on 07/03/2025)
  - Wave 6: Destatis key list of student and examination statistics, doctoral candidates statistics and guest student statistics WS 2023/2024 and SS 2024 (Federal Statistical Office, 2024): https://www.statistik-nord.de/fileadmin/Dokumente/Erhebungen/Hochschulstatis-tiken/1_Studenten_2_Pr%C3%BCfungen/Schluesselverzeichnis_WS_2023.xlsx, key 2.2 (higher education institutions by type of higher education institution and federal state) and key 2.3 (higher education institutions, numerical) (accessed on 07/03/2025)
- For variables on the (doctoral) subject:
  - Waves 1-4: Destatis subject classification WS 2017/2018 (Federal Statistical Office, 2018): https://www.statistischebibliothek.de/mir/receive/DEHeft_mods_00092410, pp. 434-436 (accessed on 04/29/2025)
  - Wave 5: Destatis subject classification WS 2021/2022 (Federal Statistical Office, 2022b): https://www.statistischebibliothek.de/mir/receive/DEHeft_mods_00143900, pp. 424-426 (accessed on 07/03/2025)
  - Wave 6: Destatis subject classification WS 2022/2023 (Federal Statistical Office, 2023a): https://www.statistischebibliothek.de/mir/receive/DEHeft_mods_00149014, pp. 79-81 (accessed on 07/03/2025)
  - Reference list of Destatis areas of study to a project-specific subject group classification, cf. appendix 10.2
- For country and nationality variables:
  - Waves 1-4: Destatis classification of states and territories (as at 04/01/2019) (Federal Statistical Office, 2019b): https://statistik.arbeitsagentur.de/DE/Statischer-Content/Grundla-gen/Klassifikationen/Staats-und-Gebietssystematik/Generische-Publikationen/2019-Staats-und-Gebietssystematik0419-pdf.pdf (accessed on 04/16/2025)
  - Waves 5-6: Destatis classification of states and territories (as at 01/01/2023) (Federal Statistical Office, 2023b): https://statistik.arbeitsagentur.de/DE/Statischer-

---

[9] In this context, it should be noted that semi-open responses were generally not coded and coding was also omitted for some other open responses. In part, they were primarily used to check which response categories might be useful for further surveys; in part, they were exclusively auxiliary variables that were intended to provide supporting information when coding other open responses. There has not yet been sufficient capacity for coding the open responses on parents' occupation in wave 1 (question E14).

  - Content/Grundlagen/Klassifikationen/Staats-und-Gebietssystematik/Generische-Publika-tionen/2023-Staats-und-Gebietssystematik-pdf.pdf (accessed on 07/03/2025)
    - NEPS-recoding map of countries (Koberg, 2016b): https://www.neps-data.de/Por-tals/0/NEPS/Datenzentrum/Forschungsdaten/SC5/6-0-0/SC5_6-0-0_Anonymisation.pdf, pp. 33-34 (accessed on 04/16/2025)
  - For income currency variables:
    - ISO 4217 Currency Codes: https://www.iso.org/iso-4217-currency-codes.html (accessed on 04/16/2025)
    - European Commission exchange rate/currency calculator: https://commission.eu-ropa.eu/funding-tenders/procedures-guidelines-tenders/information-contractors-and-ben-eficiaries/exchange-rate-inforeuro_de (accessed on 04/16/2025)

6.4   Data structure and data format

Although the data from the first to sixth survey waves were collected separately, they were merged into one data set. The cases were merged using the respondents' identification numbers assigned during the first fieldwork phase. There is a data record for each person surveyed (wide format). The order of the variables is based on the order of the corresponding questions in the questionnaire. Generated variables (cf. chapter 6.3) were positioned following the respective initial variable(s) as far as possible. The data set is provided in both Stata and SPSS format.[10]

6.5   Assignment of variable names and variable/value labels

[Variable and value labels] Short forms of the questionnaire wordings were chosen for variable and value labels. The variable and value labels are available in German and English. In Stata format, bilingual labels were stored in the same data set.[11] In SPSS format, there is a separate data set for each language.

[Variable naming] A prefix-root-suffix scheme was used for variable naming. The *prefix* contains the wave identifier using the letters a (wave 1), b (wave 2), c (wave 3), d (wave 4), e (wave 5) and f (wave 6). The *root* primarily indicates the thematic area to which the variable is assigned (cf. Table 7 for the different thematic areas and the corresponding abbreviation for the root of the variable name). The variables are also numbered within the defined thematic areas. Following that, a letter at the end of the root is used to differentiate variables that belong to the same question and thus have the same thematic differentiation and numbering (e.g. in item batteries, multiple responses or questions in which closed and open questions are combined). The *suffix* "_g#" is used to identify generated variables (cf. chapter 6.3).

---

[10] When using Stata/BE (maximum number of variables: 2,048), wave subsets of the data can be loaded using the wave prefixes in the variable names (cf. chapter 6.5), e. g. the ID variable plus all variables from the first wave via: `use pid a* using "datasetname.dta"`
[11] The command "label language en" switches to the English labels. You can switch back to the German labels using "label language de".

Thematic area abbreviation	Thematic area (English)	Thematic area (German)
ict*	informed consent	Informierte Einwilligung
dbi	basic information on doctoral phase	Eckdaten zur Promotion
did	interruption/drop-out	Promotionsunterbrechung und -abbruch
dtc	thesis topic and motives for PhD/ scientific activities	Promotionsthema und -motive/ Wissenschaftliche Aktivitäten
dcd	conditions during doctoral phase	Promotionsbedingungen
dsv	supervisor	Betreuer*in
fin	finances	Finanzierung
emp	employment	Beschäftigungssituation
dwr	daily scientific working routine	Wissenschaftlicher Arbeitsalltag
abr	international mobility	Auslandsaufenthalte/Mobilität
psy	personality/psychic traits	Persönlichkeit/psychische Eigenschaften
lcd	living conditions	Persönliche Lebenssituation
hea	health	Gesundheit
crg	career goals	Karriereziele
dem	demography	Demographie
par	social background	Soziale Herkunft
edb	educational biography	Bildungsbiographie
fed	further education	Berufsständische Weiterbildung
pop	political participation	Politische Partizipation
kom	science communication	Wissenschaftskommunikation
sum	panel commitment	Panel Commitment
inf	information	Information
cmt*	comment	Kommentar

Table 7:   Thematic areas and abbreviations for variable names in Nacaps 2018

*Variables with this root are not included in the published data.

6.6 Coding of missing values

The FDZ-DZHW has created a comprehensive system for coding missing values to ensure standardised missing coding across different DZHW data sets. Missing values are coded with three-digit negative values. The missing coding of the Nacaps data is based on this system.

The missing codes used in Nacaps 2018 are shown in Table 8 highlighted in bold. They can be assigned to three different groups.[12] In the first two groups, a distinction is made between missing values due to the respondent not answering questions (item non-response, interview break-off or panel non-

---

[12] A possible fourth group would include special missing codes that were only assigned to individual items as part of the data preparation of a specific data set. There is also a fifth group for "other missing values" such as data loss.

response) and missing values due to the filter procedure(s) or irrelevance of the question for the respondent (not applicable). The third group contains missing codes that were assigned by the primary research project or the FDZ-DZHW in the course of data preparation (edited missing value). This group also includes the coding that was assigned due to anonymization measures (cf. chapter 8) for certain variables.

Finally, it should be noted that the raw data exported from the survey system initially only distinguishes between two categories of missing values: (1) question was seen but not answered and (2) question was not seen. Missing values in the first category are coded as FDZ missing value -998 "no answer" or -988 "does not apply" during data preparation. For the second category of missing values (question not seen), a distinction must be made between whether the question was not seen because the person was not filtered into the question or because the person broke off the interview. First, during data preparation, all missing values in this category are coded as -989 "filtered". Then, the interview break-offs are coded. For this purpose, process data is used that contains information on the maximum page reached and the last page accessed in the questionnaire. This subsequent manual coding may lead to occasional discrepancies[13]. The system of missing values should therefore not be understood as a reliable technical fact.

Explicitly pointed out should be the following special cases or deviations from the system:

- For the variables cdcd37a to cdcd37g, there is the category "no answer (response category)", which could be actively selected by the respondents and which was assigned code 4 (not code -997) analogue to the questionnaire programming.
- For the variables bdwr12a to bdwr12d, cdwr19a to cdwr19d, ddwr19a to ddwr19d, edwr19a to edwr19d and fdwr19a to fdwr19d, there is the category "does not apply (response category)", which could be actively selected by the respondents and to which code 6 (not code -988) was assigned analogue to the questionnaire programming.
- For the variables cdtc07a to cdtc07d, cdtc19a, cdtc19b, ddtc07a to ddtc07d, ddtc19a, ddtc19b, edtc07a to edtc07d, edtc19a, edtc19b, fdtc07a to fdtc07d, fdtc19a and fdtc19b, there is the category "does not apply" which could be actively selected by the respondents and which was assigned code 2 (not code -988) analogue to the questionnaire programming. The addition "(response category)" in the value label was omitted here.

---

[13] It may occur that values are coded as -989 "filtered", which are actually interview break-offs, or values missing due to filtering may be incorrectly coded as -996 "interview break-off". It is therefore advisable to always consult the questionnaire for traceability of the filter management and not to evaluate interview break-offs.

Table 8:   System of the FDZ-DZHW for missing values

Value range	Code	Value label
-999 to -990: Non-response	-999	don't know
	-998	**no answer**
	-997	no answer (response category)
	-996	**interview break-off**
	-995	**not participated (panel)**
	-994	refused
-989 to -970: Not applicable	**-989**	**filtered**
	-988	**does not apply**
	-987	missing by design (questionnaire split)
	-986	missing by design (wave)[a]
	-985	missing by design (cohort)[b]
-969 to -950: Edited missing value	**-969**	**unknown missing[c]**
	-968	implausible value[d]
	-967	**anonymized**
	-966	**not determinable[e]**
	-965	invalid multiple answer
-949 to -930: Item-specific missing values	*(not assigned)*	
-929 to -920: Other missing values	-929	loss of data

[a] This value is only assigned for data sets in long format.
[b] This value is only assigned in pooled data sets.
[c] This value is assigned if no cause can be reconstructed.
[d] This value is assigned if information is categorized as implausible due to various factors.
[e] This value is assigned if clear coding is not possible, e.g. if an open response could not be coded.

7 Representativeness of the data

[Comparison to population] The adjustment of the data of the realized sample to selected variables of the population by means of weighting is a common procedure to counteract systematic selection processes and thus optimize the representativeness of the data. Due to the data situation, however, an adjustment weighting to the population is not possible, as there is no timely, unbiased official data on doctoral candidates in Germany. Although registered doctoral candidates were recorded by the Federal Statistical Office for the first time on the 1st of December 2017, the doctoral candidates were under-recorded in these statistics, as the necessary organisational and technical structures that would have enabled unbiased recording were not yet in place at that time (cf. Vollmar, 2019). The second survey as of the 1st of December 2018 also continues to show underreporting (Federal Statistical Office, 2019c). Nevertheless, in order to obtain an overview of a possible uneven distribution between the sample and the population, the distribution of people by subject group was compared between the realized sample and the data from the Federal Statistical Office (Destatis) (2019c) (Table 9).[14]

Table 9:	Distribution of doctoral candidates according to Destatis 2019 and in Nacaps 2018, 1st wave

Subject groups	Destatis 2019		Nacaps 2018, 1st wave		
	Absolute	in %	Absolute	in %	Diff.
Humanities	21.688	12,5	2.919	11,8	-0,7
Sports	1.099	0,6	165	0,7	0,1
Law, Economics and Social Sciences	30.049	17,3	4.884	19,7	2,4
Mathematics, Natural Sciences	41.810	24,1	7.024	28,3	4,2
Human Medicine, Health Sciences	36.925	21,2	3.666	14,8	-6,4
Agricultural, Forest and Nutritional Sciences, Veterinary medicine	5.158	3,0	738	3,0	0,0
Engineering Sciences	30.944	17,8	4.318	17,4	-0,4
Art, Art Theory	3.787	2,2	494	2,0	-0,2
Not specified (outside the study area structure)	2.319	1,3	578	2,3	1,0
	173.779	100,0	24.786*	100,0	

* The deviation from the values shown in chapter 5.1 is mainly due to missing values in the subject variable.

Accordingly, people doing a doctorate in law, economics and social sciences, in the subject group mathematics, natural sciences and in subjects which are not specified (outside the study area structure) are slightly overrepresented in the survey. In contrast, doctoral candidates from the human medicine, health sciences subject group are slightly underrepresented. In the other subject groups, the distributions in both data sources are almost the same.

---

[14] The first official statistics on doctoral candidates were presented in autumn 2020 (Federal Statistical Office 2020). In these statistics, the official number of doctoral candidates has increased significantly to 182,778 compared to the 2019 report, but the shifts in the proportions between the subject groups are rather small.

A comparison of the distributions by gender and subject group (Table 10) shows that women from the law, economics and social sciences subject group are slightly overrepresented in Nacaps. In contrast, both men and women are slightly overrepresented in the mathematics, natural sciences subject group. In the human medicine, health sciences subject group, both men and women are underrepresented; however, the difference is significantly greater for women than for men. There are no significant differences in the other subject groups. Also the differences in the distribution by gender between the Nacaps sample as a whole (51 percent men, 48 percent women, 1 percent diverse) and the information from official statistics (53 percent men and 47 percent women) are rather small.

Table 10:	Distribution of male and female doctoral candidates according to Destatis 2019 and in Nacaps 2018, 1st wave

Subject groups	Destatis 2019				Nacaps 2018, 1st wave			
	absolute, male	in % male	absolute, female	in % female	absolute, male*	in % male	absolute, female*	in % female
Humanities	9.541	10,3	12.147	15,0	1.009	9,1	1.595	15,2
Sports	605	0,7	494	0,6	67	0,6	85	0,8
Law, Economics and Social Sciences	15.600	16,8	14.449	17,9	2.000	18,0	2.307	21,9
Mathematics, Natural Sciences	24.230	26,1	17.380	21,5	3.398	30,6	2.747	26,1
Human medicine, Health Sciences	13.876	15,0	23.049	28,5	1.206	10,9	1.881	17,9
Agricultural, Forest and Nutritional Sciences, Veterinary medicine	1.890	2,0	3.268	4,0	194	1,7	481	4,6
Engineering Sciences	24.418	26,3	6.526	8,1	2.873	25,9	845	8,0
Art, Art Theory	1.297	1,4	2.490	3,1	114	1,0	339	3,2
Not specified (outside the study area structure)	1.221	1,3	1.098	1,4	242	2,2	248	2,4
	92.678	100,0	80.901	100,0	11.103	100,0	10.528	100,0

* The deviation from Table 9 is mainly due to missing values in the gender variable. In addition, people who stated "diverse" were not included in this analysis.

In addition, the distributions can be compared with other data sources, which themselves, however, have uncertainties regarding their representativeness. These include the statistics on completed PhD/doctorates in 2018 (Federal Statistical Office, 2019d). However, one problem with the comparison with this data source is that the discontinuation of PhD/doctorates may differ systematically between subject groups, genders or other variables. In this respect, such a comparison can be made, but is probably subject to greater uncertainty than a comparison with the statistics above.[15] A comparison with an older survey on doctoral candidates in Germany conducted by the Federal Statistical Office is also possible. This is a so-called section-7-survey, which is conducted outside of the regular statistical surveys (Federal Statistical Office, 2016). According to section 7 of the Federal Statistics Act, special surveys can be carried out (max. two times) to generate data for special purposes. As statistics on doctoral candidates did not exist until the amendment to the Higher Education Statistics Act in 2016 and the student statistics systematically underreport doctoral candidates, the section-7-survey was conducted for the first time for the 2010/11 winter semester and then again for the 2014/15 winter

---

[15] Nevertheless, if this comparison is made, doctoral candidates in the humanities (+5%) and law, economics and social sciences (+5%) are overrepresented in Nacaps, while doctoral candidates in human medicine / health sciences are underrepresented (-10%).

semester. This means that the survey for the 2014/15 winter semester is somewhat older, and the information is based on surveys of professors and doctoral candidates. In this respect, a comparison between the information on the population from this survey and Nacaps is also subject to uncertainties.[16] Another data source for comparing the distributions is the official statistics on enrolled students, which also includes enrolled doctoral candidates (Federal Statistical Office, 2019e). In this data source, however, there is a significant underreporting of doctoral candidates compared to the other two data sources on doctoral candidates, as only around 110,000 doctoral candidates are recorded here. It is therefore also to be expected that a comparison with the distribution in Nacaps is fraught with uncertainty.[17]

As there are currently no completely reliable data sources for comparing subject and gender distributions for Nacaps, no reliable assessment of the representativeness of the data can be made. For the same reason, no weighting variables can be generated. However, the statistical comparisons presented should enable data users to make an initial assessment of the representativeness of the data.

[Distribution of higher education institutions] The distribution of participating and non-participating higher education institutions was also analysed. In terms of regional distribution, there are only slight differences between the north, east, south and west regions. With a participation rate of 27.6 percent, higher education institutions from the east are slightly underrepresented, while those from the north are slightly overrepresented (41.9 percent). 32.7 percent of higher education institutions from the west and 34.9 percent from the south took part. In addition, larger higher education institutions are slightly overrepresented, while smaller universities are slightly less represented. In terms of the number of professors, 21.2 percent of smaller, 37.3 percent of medium-sized and 44.2 percent of large higher education institutions took part in Nacaps.[18] This relatively high participation rate of the large higher education institutions is probably more of an advantage for the informative value of the sample, as it significantly increases the total number of participating doctoral candidates.

[Attrition analysis] Systematic or selective panel non-response can lead to biased results. To check whether panel attrition between the survey waves occurs systematically, attrition analyses were carried out. For this purpose, five logistic regression analyses were calculated, which estimate the probability of participation in waves 2 to 6 (cf. Table 11 and Table 12). It should be noted that cases for which a request for complete deletion of data had already been made at the time of the calculations (cf. chapter 5) were not included in the calculations. At the same time, it must be taken into account that further data deletions were made after the calculations.

---

[16] When comparing these surveys, there are again differences in the humanities. This time, however, doctoral candidates from these subject groups are underrepresented in Nacaps compared to the data source used (-6 %); in law, economics and social sciences (+3 %) and human medicine / health sciences (+5 %), on the other hand, doctoral candidates are overrepresented in Nacaps.

[17] However, a comparison shows only minor deviations from the subject distribution in Nacaps. Only in the field of human medicine / health sciences are doctoral candidates overrepresented in Nacaps (+4 %).

[18] It should be noted that, given the small number of higher education institutions overall, even slight changes in participation figures have a significant impact on the percentage distribution (especially in the regional distribution). For this reason, among others, the participation rate is not reported at federal state level.

| Table 11: | Logistic regression of the probability of participation in waves 2, 3 and 4 | | | | | |

		2nd wave		3rd wave		4th wave	
Gender (ref. female)	male	0,761 (0,022)	***	0,759 (0,023)	***	0,753 (0,023)	***
	diverse	0,432 (0,097)	***	0,609 (0,137)	*	0,660 (0,147)	
	no answer	0,642 (0,221)		0,671 (0,247)		0,517 (0,216)	
Age (ref. 1991-1993)	until 1984	0,804 (0,036)	***	0,895 (0,042)	*	1,011 (0,048)	
	1985-1990	0,892 (0,030)	***	0,938 (0,033)		0,958 (0,035)	
	1994 and younger	1,099 (0,074)		1,004 (0,072)		0,979 (0,072)	
	no answer	0,276 (0,053)	***	0,358 (0,072)	***	0,337 (0,071)	***
Country of the higher education entrance qualification (ref. Germany)	abroad	0,549 (0,021)	***	0,473 (0,020)	***	0,490 (0,021)	***
	no answer	0,068 (0,028)	***	0,039 (0,020)	***	0,017 (0,013)	***
Parents' highest educational qualification (ref. no higher education degree)	PhD/doctorate	1,069 (0,044)		1,024 (0,044)		1,143 (0,049)	**
	other higher education degree	1,105 (0,033)	***	1,095 (0,034)	**	1,097 (0,035)	**
	no answer	0,470 (0,067)	***	0,575 (0,086)	***	0,611 (0,092)	**
Self-efficacy (ref. 3 "medium")	1-2 low	0,998 (0,119)		1,045 (0,131)		1,080 (0,137)	
	4-5 high	1,170 (0,052)	***	1,106 (0,051)	*	1,168 (0,055)	**
	no answer	0,382 (0,098)	***	0,501 (0,130)	**	0,539 (0,145)	*
Status of PhD/doctorate (ref. doing PhD/doctorate)	completed	2,997 (0,770)	***	2,049 (0,536)	**	2,034 (0,550)	**
	interrupted	0,896 (0,093)		0,976 (0,106)		1,052 (0,113)	
	quit	1,677 (0,487)		0,854 (0,264)		1,225 (0,376)	
	no answer	7,076 (7,861)		7,908 (8,990)		-	
Subject group (ref. Humanities)	Sports	0,989 (0,158)		0,883 (0,146)		1,035 (0,169)	
	Law, Economics and Social Sciences	0,876 (0,043)	**	0,837 (0,043)	***	0,872 (0,045)	**
	Mathematics, Natural Sciences	0,727 (0,035)	***	0,790 (0,040)	***	0,785 (0,040)	***
	Human medicine, Health Sciences	0,572 (0,033)	***	0,549 (0,033)	***	0,563 (0,034)	***
	Agricultural, Forest and Nutritional Sciences	0,660 (0,059)	***	0,724 (0,067)	***	0,720 (0,068)	***
	Engineering Sciences	0,654 (0,036)	***	0,654 (0,038)	***	0,685 (0,040)	***
	Art, Art Theory	0,913 (0,092)		0,917 (0,096)		0,882 (0,093)	
	no answer	0,757 (0,073)	**	0,803 (0,081)	*	0,724 (0,075)	**
Federal state of the doctoral higher education institute (ref. Schleswig-Holstein)	Hamburg	1,053 (0,111)		1,007 (0,111)		1,031 (0,114)	
	Lower Saxony	0,793 (0,087)	*	0,826 (0,095)		0,825 (0,095)	
	North Rhine-Westphalia	0,801 (0,074)	*	0,819 (0,079)	*	0,865 (0,083)	
	Hesse	0,794 (0,079)	*	0,864 (0,090)		0,879 (0,092)	
	Rhineland-Palatinate	0,747 (0,076)	**	0,780 (0,083)	*	0,746 (0,081)	**
	Baden-Wuerttemberg	0,967 (0,091)		0,966 (0,095)		0,991 (0,098)	
	Bavaria	0,942 (0,087)		0,938 (0,091)		0,900 (0,088)	
	Saarland	0,660 (0,078)	***	0,788 (0,097)		0,662 (0,084)	**
	Berlin	0,984 (0,102)		1,052 (0,114)		1,034 (0,112)	
	Brandenburg	1,153 (0,134)		1,087 (0,131)		1,140 (0,138)	
	Saxony	1,010 (0,113)		1,088 (0,127)		1,022 (0,120)	
	Saxony-Anhalt	0,791 (0,109)		0,879 (0,127)		0,750 (0,112)	
	Thuringia	0,751 (0,087)	*	0,830 (0,101)		0,936 (0,114)	
	no answer	-		-		-	
employed in science (ref. no)	yes	1,255 (0,039)	***	1,274 (0,042)	***	1,250 (0,042)	***
	no answer	0,476 (0,191)		0,463 (0,211)		0,379 (0,204)	
	Pseudo-R²	0,136		0,118		0,113	
	N	28.349		28.349		28.147	

Notes: Odds ratios, standard errors in parentheses; *** p<0.001, ** p<0.01, * p<0.05

Table 12:	Logistic regression of the probability of participation in waves 5 and 6				

		5th wave		**6th wave**	
Gender (ref. female)	male	0,738 (0,024)	***	0,812 (0,027)	***
	diverse	0,635 (0,157)		0,421 (0,126)	**
	no answer	0,735 (0,303)		0,808 (0,343)	
Age (ref. 1991-1993)	until 1984	1,054 (0,054)		1,084 (0,057)	
	1985-1990	1,041 (0,041)		1,062 (0,043)	
	1994 and younger	0,851 (0,070)	*	0,828 (0,070)	*
	no answer	0,413 (0,093)	***	0,351 (0,089)	***
Country of the higher education entrance qualification (ref. Germany)	abroad	0,457 (0,022)	***	0,461 (0,023)	***
	no answer	0,030 (0,020)	***	0,048 (0,029)	***
Parents' highest educational qualification (ref. no higher education degree)	PhD/doctorate	1,159 (0,054)	**	1,203 (0,057)	***
	other higher education degree	1,178 (0,040)	***	1,165 (0,041)	***
	no answer	0,550 (0,097)	***	0,584 (0,106)	**
Self-efficacy (ref. 3 "medium")	1-2 low	1,021 (0,142)		1,014 (0,148)	
	4-5 high	1,153 (0,059)	**	1,188 (0,063)	**
	no answer	0,493 (0,150)	*	0,591 (0,178)	
Status of PhD/doctorate (ref. doing PhD/doctorate)	completed	2,042 (0,623)	*	1,900 (0,573)	*
	interrupted	0,917 (0,108)		0,952 (0,115)	
	quit	1,202 (0,418)		1,038 (0,362)	
	no answer	-		-	
Subject group (ref. Humanities)	Sports	0,824 (0,146)		0,862 (0,156)	
	Law, Economics and Social Sciences	0,851 (0,046)	**	0,861 (0,048)	**
	Mathematics, Natural Sciences	0,767 (0,041)	***	0,823 (0,046)	***
	Human medicine, Health Sciences	0,570 (0,037)	***	0,630 (0,042)	***
	Agricultural, Forest and Nutritional Sciences	0,561 (0,059)	***	0,628 (0,067)	***
	Engineering Sciences	0,667 (0,042)	***	0,604 (0,039)	***
	Art, Art Theory	0,722 (0,083)	**	0,849 (0,098)	
	no answer	0,805 (0,088)	*	0,694 (0,082)	**
Federal state of the doctoral higher education institute (ref. Schleswig-Holstein)	Hamburg	1,005 (0,119)		0,892 (0,106)	
	Lower Saxony	0,894 (0,110)		0,808 (0,100)	
	North Rhine-Westphalia	0,861 (0,089)		0,740 (0,077)	**
	Hesse	0,839 (0,094)		0,706 (0,080)	**
	Rhineland-Palatinate	0,736 (0,086)	**	0,645 (0,076)	***
	Baden-Wuerttemberg	0,994 (0,106)		0,914 (0,097)	
	Bavaria	0,912 (0,096)		0,834 (0,087)	
	Saarland	0,738 (0,100)	*	0,642 (0,089)	**
	Berlin	1,054 (0,123)		0,922 (0,108)	
	Brandenburg	1,151 (0,149)		0,916 (0,120)	
	Saxony	1,177 (0,147)		0,960 (0,121)	
	Saxony-Anhalt	0,768 (0,124)		0,779 (0,126)	
	Thuringia	0,832 (0,110)		0,707 (0,095)	**
	no answer	-		-	
employed in science (ref. no)	yes	1,155 (0,041)	***	1,242 (0,046)	***
	no answer	0,320 (0,197)		0,217 (0,161)	*
	Pseudo-R²	0,100		0,096	
	N	28.134		28.134	

Notes: Odds ratios, standard errors in parentheses; *** p<0.001, ** p<0.01, * p<0.05

In order to investigate the extent to which the probability of participation in the follow-up waves is related to the characteristics of the respondents (MAR), the explanatory models contain demographic variables (gender, age, country of higher education entrance qualification, parents' highest educational qualification), personality traits (self-efficacy) and contextual variables of the PhD/doctorate (status of PhD/doctorate, doctoral subject (subject group), federal state of the doctoral higher education institution, employment in science during the doctorate) as independent variables, which were collected as part of the initial survey and recorded for all status groups. Missing values on the explanatory variables were each considered as a separate category in the analyses[19]. This shows that unit nonresponse on variables in the model from the initial survey is largely a strong predictor of panel attrition in the follow-up surveys.

Compared to female respondents, male respondents were significantly less likely to take part in the follow-up surveys. People with the gender identity "diverse" were also less likely to participate in the second, third and sixth wave. Age also had an influence on participation behaviour. People born before 1991 were significantly less likely to participate in the second survey; and a negative effect can also be shown for the third wave for the oldest category of those born up to and including 1984. People born in 1994 or later are less likely to participate in waves 5 and 6. A higher education entrance qualification obtained abroad significantly reduces the probability of participation in all five follow-up surveys. Compared to respondents from non-academic parental homes, a higher education degree from one or both parents favours the probability of participation in all five follow-up surveys. A positive effect on the probability of participating from the fourth wave onwards can also be demonstrated if at least one parent has completed a PhD/doctorate. In addition, a high expectation of self-efficacy increases the probability of participating in the follow-up surveys.

Respondents who had already completed their PhD/doctorate at the time of the initial survey were significantly more likely to take part in the follow-up surveys than doctoral candidates. An interrupted or discontinued PhD/doctorate, on the other hand, has no effect on the probability of participation. At the same time, subject effects can be observed. Compared to the reference category of respondents from the subject group of humanities, a significantly negative effect on the probability of participation in the follow-up waves can be demonstrated for all subject groups – with the exception of the subject groups of sports and art / art theory (except in wave 5). Significant differences in participation in the follow-up waves can also be found in the federal states in which the higher education institutions are located. Compared to Schleswig-Holstein, which is the reference category here, significant negative effects can be found for Lower Saxony (second wave), North Rhine-Westphalia (second, third & sixth wave), Hesse (second & sixth wave), Rhineland-Palatinate (all waves), Saarland (second, fourth, fifth & sixth wave) and Thuringia (second & sixth wave). Finally, the field of employment during the PhD/doctorate also has an influence on participation in the follow-up surveys. Being employed in science during the PhD/doctorate favours participation in the Nacaps follow-up surveys.

---

[19] Cases with missing values on the variable "status of PhD/doctorate" in the models for estimating the probability of participation in the fourth, fifth and sixth wave are an exception. All respondents with missing values on this variable did not participate in the respective survey, so Stata excluded these 202 cases from the analyses.

8   Anonymization

[Legal Framework] The EU General Data Protection Regulation (GDPR) and the Federal Data Protection Act (BDSG) in its revised version of 30 June 2017 apply to personal data[20] collected by the DZHW in voluntary surveys.[21] Accordingly, personal data for secondary scientific use (without a declaration of consent for secondary use of the personal data) must be processed "in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person" (Art. 4 para. 5 GDPR; see also Art. 89 GDPR and Recital 26 GDPR).

[Ensuring data protection at the FDZ-DZHW] At the FDZ-DZHW, the data protection of the respondents is ensured by a combination of legal-organisational, technical and statistical measures.[22] The combinations result in four potential access ways, which are summarized in Table 13.[23]

Table 13:   Access ways in the FDZ-DZHW

	Campus Use File (CUF): Download	Scientific Use File (SUF): Download	Scientific Use File (SUF): Remote Desktop	Scientific Use File (SUF): On-Site
legal-organisational measures	very low	low	moderate	high
technical measures	very low	low	moderate	high
statistical measures	very high	high	moderate	low

The more data access is legally-organisationally and technically controlled and restricted, the lower the risk of de-anonymization of the data. As a result, the data must be anonymized less by means of statistical measures, i.e. less information has to be removed from it – which means the greater its analysis potential remains.

For the Nacaps 2018 data, all access ways are available. The specific statistical anonymization measures carried out are explained in more detail below.

[Statistical anonymization measures] As part of the statistical anonymization measures, all information was first checked to see whether it could be used to directly identify individuals. These *direct*

---

[20] "'Personal data' means any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person" (Art. 4 para. 1 GDPR).

[21] The GDPR generally applies within the EU and therefore also to the DZHW. The BDSG in its new version of 30 June 2017 (Act to Adapt Data Protection Law to the Regulation (EU) 2016/679 and to Implement Directive (EU) 2016/680, (Data Protection Adaptation and Implementation Act-EU (DSAnpUG-EU)) also applies in part, as the DZHW GmbH is legally considered a public body of the federal government (sect. 2 para. 3 BDSG). The federal government holds the absolute majority of shares in DZHW GmbH and the institute fulfils public administration tasks for the federal government in the broadest sense.

[22] The FDZ's data protection concept is based on the portfolio approach of Lane, Heus and Mulcahy (2008, p. 6ff.), which is already used by the Leibniz Institute for Educational Trajectories (LIfBi) (cf. Koberg, 2016a, p. 699ff.) and the RDC of the German Federal Employment Agency at the Institute for Employment Research (cf. Hochfellner, Müller, Schmucker, & Roß, 2012, p. 9f.).

[23] For more detailed information, see https://www.fdz.dzhw.eu/en/data-usage

identifiers, such as names, addresses and email addresses, were already recorded in a separate data set during the fieldwork phase and are therefore not included in the data per se. The original identification number was removed and replaced by a new randomly assigned identification number.

In addition, Ebel and Meyermann's recommendation to delete open responses was followed even if the questions themselves are unproblematic. This is because there is a risk that study participants may have disclosed critical information in actually harmless questions with open response option that could lead to identification (cf. Ebel & Meyermann, 2015, p. 5).

After that, quasi-identifiers were determined, i.e. information that, in combination or by adding external information, is suitable for indirectly identifying a person. In Nacaps 2018 this includes, for example, specific information on key data about doctoral phase (e. g. university, subject or grade) or the progress of the doctorate (e. g. date of completion), doctoral programmes and doctoral funding, doctoral supervisors, stays abroad and scientific activities (e.g. publications and applications), employment (e. g. sector, occupational status, place of work) and further training, educational stages prior to the doctorate, sociodemographics (e. g. year of birth, country of birth/nationality or information about parents) and personal life situation (e. g. children, main residence). To prevent unambiguous assignment of the data to a specific person, it was determined for these key characteristics – depending on the access way – whether they may only be available in an aggregated form or not at all.

Finally, the data was checked for sensitive information, e.g. on health, sexual orientation and political attitudes. Although this information is not necessarily suitable for re-identification of individuals or institutions, it does pose a potentially increased risk of harm to respondents in the event of de-anonymization (cf. Koberg, 2016a, p. 694) and is therefore particularly worthy of protection (Art. 9 GDPR, Recital 51 GDPR). In Nacaps 2018 this includes, for example, information on health, political participation, conflicts, discrimination, sexual harassment and bullying. For the sensitive information, it was also determined – depending on the access way – whether it may only be available in an aggregated form or not at all.

The dataset report shows which variables are contained in the data provided and via which access way(s) they can be used.

For the access way "Campus Use File (CUF): Download", in addition to the information reduction at variable level, a randomly selected subsample of the data (25 % of respondents) was drawn, i.e. the number of cases was also reduced.

9 Literature

AAPOR (2016). Standard Definitions. Final Dispositions of Case Codes and Outcome Rates for Surveys. https://aapor.org/wp-content/uploads/2022/11/Standard-Definitions20169theditionfinal.pdf (accessed on: 11/18/2025).

Brandtstädter, J., & Renner, G. (1990). Tenacious goal pursuit and flexible goal adjustment: Explication and age-related analysis of assimilative and accommodative strategies of coping. Psychology and Aging, 5(1), pp. 58–67. doi: 10.1037/0882-7974.5.1.58.

Ebel, T., & Meyermann, A. (2015). Hinweise zur Anonymisierung von quantitativen Daten. Forschungsdaten Bildung informiert. Bd.3. Frankfurt am Main: Deutsches Institut für Internationale Pädagogische Forschung. doi: 10.25656/01:21970.

Hochfellner, D., Müller, D., Schmucker, A., & Roß, E. (2012). Datenschutz am Forschungsdatenzentrum. (FDZ-Methodenreport Nr. 6). Nürnberg: Institut für Arbeitsmarkt- und Berufsforschung (IAB).

Koberg, T. (2016a). Disclosing the National Educational Panel Study. In H.-P. Blossfeld, J. v. Maurice, M. Bayer, & J. Skopek (eds.), Methodological Issues of Longitudinal Surveys. The example of the National Educational Panel Study (pp. 691–708). Wiesbaden: Springer VS. doi: 10.1007/978-3-658-11994-2_38.

Koberg, T. (2016b). Starting Cohort 5: First-Year Students (SC5). SUF Version 6.0.0. Anonymization Procedures. https://www.neps-data.de/Portals/0/NEPS/Datenzentrum/Forschungsdaten/SC5/6-0-0/SC5_6-0-0_Anonymisation.pdf (accessed on: 06/05/2020).

Lane, J., Heus, P., & Mulcahy, T. (2008). Data access in a cyber world: Making use of cyberinfrastructure. Transactions on Data Privacy, 1(1), 2–16.

Lenzner, T., Neuert, C., & Otto, W. (2015). Kognitives Pretesting. Mannheim: GESIS – Leibniz-Institut für Sozialwissenschaften (GESIS Survey Guidlines). doi: 10.15465/gesis-sg_010.

Prüfer, P., & Rexroth, M. (2000). Zwei-Phasen-Pretesting. (ZUMA-Arbeitsbericht 2000/08). Mannheim: Zentrum für Umfragen, Methoden und Analysen (ZUMA). https://nbn-resolving.org/urn:nbn:de:0168-ssoar-200864 (accessed on: 04/16/2025).

Rußmann, M., Netz, N., & Schwabe, U. (2025). Doctoral Students' Life Satisfaction Throughout the Covid-19 Pandemic: Inequalities by Parenthood and Gender. Social Indicators Research. doi: 10.1007/s11205-025-03648-0.

Federal Statistical Office (Statistisches Bundesamt) (2016). Promovierende in Deutschland. Wintersemester 2014/2015. Wiesbaden: Statistisches Bundesamt.

Federal Statistical Office (Statistisches Bundesamt) (2018). Bildung und Kultur. Studierende an Hochschulen. Wintersemester 2017/2018. (Fachserie 11/4.1). Fächersystematik (Fächergruppen, Studienbereiche und Studienfächer). Wiesbaden: Statistisches Bundesamt.

Federal Statistical Office (Statistisches Bundesamt) (2019a). Schlüsselverzeichnisse für die Studenten- und Prüfungsstatistik, Promovierendenstatistik WS 2018/2019 und SS 2019. Wiesbaden: Statistisches Bundesamt.

Federal Statistical Office (Statistisches Bundesamt) (2019b). Staats- und Gebietssystematik. Stand 01.04.2019. Wiesbaden: Statistisches Bundesamt.

Federal Statistical Office (Statistisches Bundesamt) (2019c). Bildung und Kultur. Promovierendenstatistik: Analyse zu Vollständigkeit und Qualität der zweiten Erhebung 2018. Wiesbaden: Statistisches Bundesamt.

Federal Statistical Office (Statistisches Bundesamt) (2019d). Bildung und Kultur. Prüfungen an Hochschulen 2018. (Fachserie 11/4.2). Wiesbaden: Statistisches Bundesamt.

Federal Statistical Office (Statistisches Bundesamt) (2019e). Bildung und Kultur. Studierende an Hochschulen. Wintersemester 2018/19. (Fachserie 11/4.1). Wiesbaden: Statistisches Bundesamt.

Federal Statistical Office (Statistisches Bundesamt) (2020). Statistik der Promovierenden 2019. Wiesbaden: Statistisches Bundesamt.

Federal Statistical Office (Statistisches Bundesamt) (2022a). Schlüsselverzeichnisse für die Studenten- und Prüfungsstatistik, Promovierendenstatistik und Gasthörerstatistik WS 2021/2022 und SS 2022. Wiesbaden: Statistisches Bundesamt.

Federal Statistical Office (Statistisches Bundesamt) (2022b). Bildung und Kultur. Studierende an Hochschulen. Wintersemester 2021/2022. (Fachserie 11/4.1). Fächersystematik (Fächergruppen, Studienbereiche und Studienfächer). Wiesbaden: Statistisches Bundesamt.

Federal Statistical Office (Statistisches Bundesamt) (2023a). Bildung und Kultur. Studierende an Hochschulen. Wintersemester 2022/2023. (Fachserie 11/4.1). Fächersystematik (Fächergruppen, Studienbereiche und Studienfächer). Wiesbaden: Statistisches Bundesamt.

Federal Statistical Office (Statistisches Bundesamt) (2023b). Staats- und Gebietssystematik. Stand 01.01.2023. Wiesbaden: Statistisches Bundesamt.

Federal Statistical Office (Statistisches Bundesamt) (2024). Schlüsselverzeichnisse für die Studenten- und Prüfungsstatistik, Promovierendenstatistik und Gasthörerstatistik WS 2023/2024 und SS 2024. Wiesbaden: Statistisches Bundesamt.

Vollmar, M. (2019). Neue Promovierendenstatistik: Analyse der ersten Erhebung 2017. WISTA Wirtschaft und Statistik, 1, 68-79.

Weichbold, M. (2019). Pretest. In N. Bauer, & J. Blasius (eds.), Handbuch Methoden der empirischen Sozialforschung (pp. 349-356). Wiesbaden: Springer VS.

Wegner, A., & Briedis, K. (2020). National Academics Panel Study – Die Längsschnittstudie zu Promovierenden und Promovierten in Deutschland. Wissenschaftliches Projektkonzept. https://www.nacaps.de/files/downloads/nacaps_wisskonzept.pdf (accessed on: 04/16/2025).

10 Appendix

10.1 Documentation of the origin of secondarily used questions

Due to its extent, the documentation of the origin of secondarily used questions is provided in a separate file. This can be accessed via the following link:

https://metadata.fdz.dzhw.eu/public/files/data-packages/stu-nac2018$-3.0.0/attachments/nac2018_QuestionOrigin.xlsx

10.2 Reference list: Destatis areas of study to project-specific subject group classification

Table 14:	Allocation of Destatis areas of study to project-specific subject group classification

DZHW subject group (code)	DZHW subject group (label)	Destatis area of study (code)
1	agricultural, forest and nutritional sciences, veterinary medicine	51; 57; 58; 59; 60
2	architecture and civil engineering	66; 68
3	biology	42
4	chemistry	40
5	electrical engineering	64
6	German studies	9
7	history	5
8	computer science	71
9	art, art theory	74; 75; 76; 77; 78
10	mechanical engineering	63
11	mathematics	37
12	medicine	48; 49; 50
13	physics	39
14	political and social sciences	23; 24; 25; 26; 27; 29; 34
15	psychology and educational sciences	32; 33
16	law	28
17	other engineering sciences	61; 62; 65; 67; 69; 70; 72
18	other natural sciences	36; 41; 43; 44
19	linguistic and cultural studies	1; 2; 3; 4; 6; 7; 8; 10; 11; 12; 13; 14; 18; 19
20	business and economics	30; 31
21	sports	22
-966	not determinable	83