ERiK report also available via **www.dji.de/erik**

ERiK

Methodological Report IV

Implementation and Data Quality of the ERiK-Surveys 2022

Forschungsverbund
tu DJI
Deutsches Jugendinstitut
Technische Universität Dortmund

DJI Deutsches Jugendinstitut

wbv

ERiK-Methodological Report IV

Implementation and Data Quality of the ERiK-Surveys 2022

ERiK

Methodological Report IV

Implementation and Data Quality of the ERiK-Surveys 2022

Forschungsverbund
tu DJI
Deutsches Jugendinstitut
Technische Universität Dortmund

DJI
Deutsches
Jugendinstitut

wbv

Researching children, youth and families at the intersection of science, policy, and professional practice

The German Youth Institute (DJI) is one of the largest social science research institutes in Europe with an experience of over 60 years. The DJI conducts empirical studies into the life situations of children, young people and families, and provides policy advice to the German national government, the German federal states and local authorities as well as key impulses for professional practice.

Founded in 1963, the governing body of the institute is a non-profit association with members from the fields of politics and science as well as from child, youth and family welfare institutions and organisations. The DJI receives the majority of its funding from the Federal Ministry for Family Affairs, Senior Citizens, Women and Youth, and the German federal states. The Federal Ministry of Education and Research, the European Commission as well as foundations and other science funding institutions provide additional funding.

Currently, about 500 staff members work and conduct research at the two locations in Munich and Halle (Saale).

Authors:
Sonja Herrmann, Jakob J. Gilg, Franz L. Classe, Diana D. Schacht, Benjamin Gedon, Susanne Kuger

Design and layout of inner section:
Christiane Zay, Passau

Interior typesetting:
paginamedia

Editing:
Baker & Company

2024 wbv publication
a division of
wbv Media GmbH & Co. KG, Bielefeld

Overall production:
wbv Media, Bielefeld,
Auf dem Esch 4, 33619 Bielefeld,
service@wbv.de
wbv.de

Cover design:
Christiane Zay, Passau

Order number: I77840
ISBN Print: 9783763977840
ISBN E-Book: 9783763977857
DOI: 10.3278/9783763977857

Printed in Germany

Bibliographic information of the German National Library
The German National Library lists this publication in the German National Bibliography; detailed bibliographic data is available on the Internet at http://dnb.d-nb.de.

Funded by:

Federal Ministry for
Family Affairs, Senior Citizens,
Women and Youth

Contents

1. Preface

The 'ERiK Methodological Report IV' (in German: ERiK-Methodenbericht IV) is the fourth methodological report in the study 'An indicator-based monitoring of structural quality in the German early childhood education and care system' (in German: Entwicklung von Rahmenbedingungen in der Kindertagesbetreuung – indikatorengestützte Qualitätsbeobachtung – ERiK). ERiK is a national study that aims to provide data-based knowledge on the quality of early childhood education and care (ECEC) in Germany. The study includes regular surveys of parents, directors of day-care centres, pedagogical staff, family day-care workers, youth welfare offices and providers of day-care centres. The DJI conducted the first surveys in 2020. In 2022, further surveys were conducted with the aforementioned target groups as well as an additional survey of 4 to 6-year-old children who were cared for in a day-care centre in 2020 and 2021.

The 'ERiK Methodological Report III' (Schacht et al. 2023) thoroughly examines the target population and sampling strategies employed in the ERiK-Surveys 2022. By contrast, this report focuses on implementing the sample and survey designs and assessing the data quality of the ERiK-Surveys 2022. Taken together, these two methodological reports encompass comprehensive background information on the ERiK-Surveys 2022.

The information presented in 'ERiK Methodological Report IV' is a collective result of the collaborative efforts of the entire ERiK team, who actively contributed to the ERiK-Surveys 2022 at the German Youth Institute (Deutsches Jugendinstitut – DJI). Special acknowledgement is extended to the members of the ERiK steering group, namely Prof. Dr. Bernhard Kalicki, Dr. Sina Fackler, Dr. Christiane Meiner-Teubner and Christine Bopp. Additionally, gratitude is expressed to the ERiK team colleagues for their valuable feedback, including Janette Buchmann, Johanna Romefort, Dr. Theresia Pachner, Melina Preuß, Lisa Ulrich and Nadira Tursun.

The monitoring project contributes to an ongoing effort to advance the quality of early childhood education and care in Germany. As a result, the publication is enriched by the diverse discussions held with colleagues of the Federal Ministry for Family Affairs, Senior Citizens, Women and Youth (BMFSFJ), representatives from the ministries of the federal states, and experts from other fields. Engaging with different committees and stakeholders has been instrumental in gathering valuable technical insights for the report. In addition, we would like to thank Dr. Christian Nagel from the department for Data & Innovation at the City of Munich (Landeshauptstadt München) for his statistical expertise. We would also like to thank Dr. Jean Philippe Décieux of the Federal Institute for Population Research (BiB) for reviewing the report and making helpful suggestions for improvement.

Furthermore, we would like to express our gratitude to the infas Institute for Applied Social Sciences and the SOKO Institute for Social Research and Communication for their dedication and efforts in conducting the ERiK-Surveys 2022.

In addition, the report has been enriched by the invaluable practical assistance of numerous individuals. We extend our special appreciation to Aydan Mammadova for her organisational support. Furthermore, our gratitude extends to Isabel Becker, Eugenia Zimmermann, Gitta Metzger, and the dedicated student assistants within the ERiK team.

Munich, September 2024

Dr. Sonja Herrmann – *Deputy Head of the ERiK Project, DJI*

Jakob J. Gilg – *Research Associate within the Department of Social Monitoring and Methodology, DJI*

Franz L. Classe – *Research Associate within the Department of Social Monitoring and Methodology, DJI*

Dr. Diana D. Schacht – *Senior Quantitative Researcher, DJI*

Benjamin Gedon – *Research Associate, SOS Kinderdorf*

Prof. Dr. Susanne Kuger – *Research director of the German Youth Institute (DJI) and Full Professor for Empirical Social and Education Research in Childhood and Adolescence at LMU Munich*

2. Introduction

The study entitled 'An indicator-based monitoring of structural quality in the German early childhood education and care system' (ERiK) is a nationwide study with the objective of offering data-driven insights into the quality of early childhood education and care (ECEC) in Germany. In 2020, the DJI conducted the initial surveys targeting directors of day-care centres, pedagogical staff, family day-care workers, youth welfare offices, and childcare centre providers. In 2022, the surveys were repeated with the previously mentioned groups and supplemented by an additional survey focussing on 4 to 6-year-old children who received childcare in 2020 and 2021. Furthermore, the ERiK parent survey is interconnected with the DJI Childcare Study (KiBS). KiBS was extended by an ERiK module comprising additional questions about the quality of day-care (for more information on this topic please see Section *7. ERiK Parent Survey* in Schacht et al. (2023); for more information on KiBS in general please see Wieschke/Lippert/Kuger (2023)). Further surveys are planned for directors of day-care centres, pedagogical staff, family day-care workers, youth welfare offices, and childcare centre providers in 2024.

This report introduces the ERiK-Surveys 2022 and focuses on the following aspects of the surveys:

1. Implementation of the sample and survey designs (see Chapter 3),
2. Evaluation of the data quality (see Chapter 4)

The chapter 'Implementation of the sample and survey design' delineates the field measures, field progress and modifications of the originally planned fieldwork. The subsequent chapter 'Evaluation of the data Quality' appraises data selectivity and modifications of the target population, outlines the weighting procedure, and evaluates the information content of the data at the federal and state levels. The report concludes with a recapitulation and recommendations for data users. For in-depth discussions regarding the target population and sampling strategies of the ERiK-Surveys 2022, readers can refer to the 'ERiK Methodological Report III' (Schacht et al., 2023). The fieldwork documents, including invitation and reminder letters, data protection information, and questionnaires, are accessible in an online appendix to this report (www.dji.de/erik/MBIV_Anhang).

3. Implementation of the ERiK-Surveys 2022

The implementation of the ERiK-Surveys 2022 involved developing questionnaires, organising specific fieldwork interventions for target groups, and conducting fieldwork. For detailed information on questionnaire revisions and the planned fieldwork interventions please see 'ERiK Methodological Report III' (Schacht et al. 2023). The following section outlines details on changes in the fieldwork process compared to the planned design, the response rates achieved, mode choices in all ERiK surveys and the regional coverage. The chapter concludes with an overview of the obstacles hindering potential respondents' participation based on the ERiK Nonresponse-Surveys 2022.

3.1 Modifications from Planned Fieldwork Efforts and Contacts

The planned fieldwork efforts and contacting strategies of the ERiK-Surveys 2022 were outlined in the ERiK Methodological Report III (Schacht et al. 2023). During the field period, some variations from the planned design became necessary and were thus implemented. These changes in the implementation of the different ERiK-Surveys 2022 are described in the following sections.

Directors and Pedagogical Staff

The planned implementation of the ERiK-Surveys of directors and pedagogical staff (as described in Section 6.1 in Schacht et al. 2023) was mostly adhered to with the field period starting on 2 February 2022. The reminders were delayed with the reminder letter coming on 1 March 2022 – two weeks later than originally planned, and the reminder by phone in conjunction with the ERiK-Nonresponse Survey between 09 March 2022 and 14 April 2022 – one week later than planned. Consequently, the field period was extended by one week until 25 April 2022.

Two minor changes were made to the online questionnaires during the field period. Firstly, the ERiK team added a confirmation question to the survey of pedagogical staff (in addition to the already existing screening question) to more precisely narrow down the target population by excluding members of teams of directors. Secondly, in the survey of directors, a more precise note was added to the question on costs for parents. The new note specified that respondents should state the *monthly* costs *per child*. These two changes in the surveys were implemented on 8 February 2022.

During the field preparation of the questionnaire for pedagogical staff, a change was made to the layout of the question on the respondents' gender in the paper questionnaire for pedagogical staff to save space. In 2022 the items were arranged horizontally next to each other instead of underneath each other as in 2020 (see Figure A.0-1 in the Appendix). This change caused confusion and misattribution in the questionnaire because the box for 'männlich' (male) was closer to the label 'weiblich' (female). This was visible in the results, where 26 % of total respondents indicated being male, while the official KJH statistics[1] suggest that the true population percentage of males is only about 7 % (Research Data Centre of the Statistical Offices of the Federal States 2022). Due to this severe deviation, the ERiK team decided to mark this variable for all paper questionnaires as implausible (.p) in the data set. This means that n=1,643 cases (32 % of the paper sample) were set to implausible in the ERiK-Surveys 2022.

1 The child and youth welfare statistics, KJH statistics for short, include: 1. data on various forms of educational assistance and administrative tasks of the youth welfare offices, 2. data on youth work measures of public and independent providers, 3. data on childcare in various youth welfare facilities and the persons working there as well as on publicly funded family day-care for children, 4. data on the expenditure and income for child and youth welfare (translated from Research Data Centre of the Statistical Offices of the Federal States 2024).

Due to the concurrent mixed mode design that was implemented in the ERiK-Surveys 2022, some directors participated twice – once online and once on the paper questionnaire. In these instances, the more complete questionnaires were retained, while the duplicates were removed from the net data set.

Youth Offices and Family Day-Care Workers

The surveys of youth offices and family day-care workers were largely implemented as intended. The youth offices received the advance letters and were then contacted via telephone between 12 January and 17 January 2022. Afterwards, the youth offices received their survey materials and the materials to forward to family day-care workers according to the contact options they were sampled for (see Sections 5.2 and 6.2 in Schacht et al. 2023). The initial reminders were sent as scheduled, two weeks after the start of the field period, and consisted of letters to 514 youth offices and 27 e-mails to youth offices. The family day-care workers in 139 youth offices were reminded via paper materials from the youth offices, while 413 youth offices sent e-mail reminders to their family day-care workers.

Due to relatively low response rates from youth offices, it was decided to add a reminder via telephone which reached 88 youth offices between 28 February and 04 March.

On 15 March 2022, the second reminders were sent via e-mail as planned. The reminders were sent to all youth offices that had not actively refused to participate in the survey, and a total of 529 youth offices were reached.[2]

A second telephone reminder was added to increase the youth office response rate. From 24 March onwards, youth offices that had not yet fully participated in the survey and for which there had been no response from family day-care workers (n=208) were contacted by telephone again.

To account for the added fieldwork efforts, the field period was extended by one week until 8 April 2022.

After carrying out the ERiK-Survey 2022 of youth offices, it came to our attention that two youth welfare office districts had merged into one, thus reducing the total number of youth offices relevant for our surveys from 569 to 568. This was the the district of 'Eisenach, Stadt' in Thuringia which merged with 'Wartburgkreis' on 1 January 2021.

Providers of Day-Care Centres

The 2024 ERiK-Survey of providers was implemented as planned and ran from 03 January 2022 to 31 March 2022, with only two minor additions. Firstly, on 07 March 2022 an update was made to the filtering in the online instrument. Anyone who did not answer the first screening question to determine their membership to the target population was asked the confirmation question to determine their eligibility. Secondly, to increase the response rates in certain federal states with low response rates, the ERiK team decided to add a third reminder in these states. These e-mail reminders were sent on 24 March 2022 to providers in Berlin, Bremen, Hamburg, Brandenburg, Mecklenburg-Western Pomerania, Rhineland-Palatinate, Saarland, Saxony, Saxony-Anhalt, Schleswig-Holstein and Thuringia.

During the field period, a provider of a day-care centre contacted the survey institute, expressing interest in participating. Although the provider was not initially included in our address list, we allowed them (and potentially others with similar requests) to participate in order to achieve a complete population survey. The provider received the survey materials via e-mail. However, this provider from Bavaria was the only one who contacted the ERiK team or survey institute.

Children

The structure of the ERiK-Survey of children was as planned (see Section 6.4 in Schacht et al. 2023), although the timeline was partly adjusted. At the beginning of March 2022, the day-care centres were contacted, and the telephone reminders commenced in mid-March, as scheduled. However, the reminders were extended by one week until 06 April 2022. The survey of parents concluded in accordance with the original plan on 18 April. The child interviews were conducted in the centres from May until the beginning of September 2022, rather than concluding in July, as initially proposed. It became evident during

2 This number includes youth offices that already participated in their survey if not all of their family day-care workers also responded, as they should be reminded as well.

the fieldwork period that it would be unfeasible to survey a sufficient number of centres with four children each within the planned time frame. It was therefore decided that, in addition to extending the field period, the four target children would be interviewed, as previously planned, but that the two replacement children per centre would also be included. This resulted in a maximum of six children being interviewed per centre, rather than four, and thus a slightly higher clustering of children.[3] For more details on the ERiK-Survey of children see the infas methodological report (von der Burg et al. 2021, (in German)).

3.2 Fieldwork

Figures 3.2-1, 3.2-2 and 3.2-3 show the weekly number of questionnaires completed in 2022 separately for each target population. Fieldwork progress was monitored to enable us to intervene in case of unexpectedly low response rates. The figures also show the time of the fieldwork measures carried out, namely the sending of the reminder letters or the start/end of the field periods in which the telephone reminders were carried out.

For all target populations, there is only one tranche in the ERiK-Surveys 2022, with the result that for each of the target populations, the entire sample was contacted at the beginning of the field time. This is a major difference to the ERiK-Surveys 2020, where the samples of directors, pedagogical staff in child day-care centres as well as the population of providers were contacted in two separate phases. This had been done to allow for adjustments in the contacting procedure. Based on the experience gained in 2020, this was no longer considered necessary for the ERiK-Surveys 2022 (Schacht et al. 2023).

Directors and pedagogical staff mostly answered the questionnaires in the first two months of fieldwork

Figure 3.2-1 shows that the first questionnaires were filled out by the directors and pedagogical staff directly after the fieldwork started at the beginning of February. In 2022, the largest proportion of questionnaires sent to pedagogical staff and directors were answered within the first two months of the survey. The despatch of the reminder letters at the beginning of March and the initiation of the telephone reminder seem to have had a positive impact on the number of completed questionnaires. This is a similar pattern to the ERiK-Surveys 2020.

About 39 % of the directors' questionnaires and about 33 % of the pedagogical staff questionnaires were answered after the telephone reminders, which were conducted between 9 March 2022 and 14 April 2022 (see Section 3.6). This is considerably more than in the ERiK-Surveys 2020.

Most of the youth welfare offices and family day-care workers completed the questionnaires in the first two months of the field period

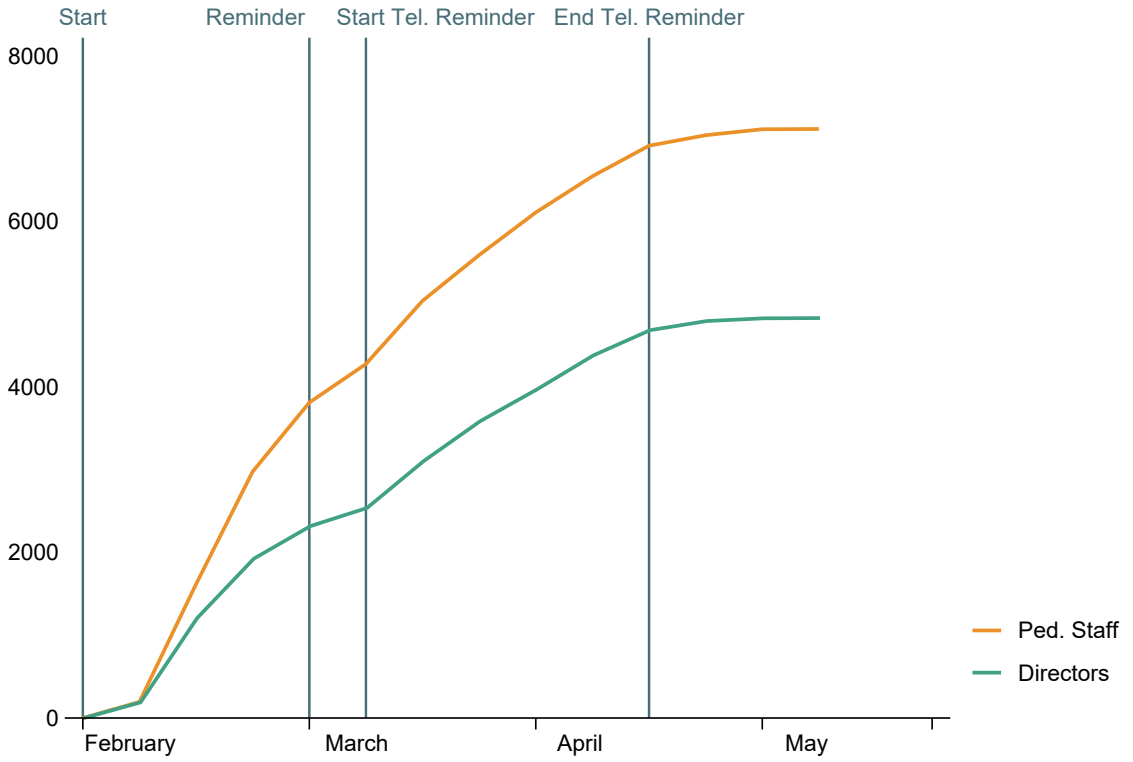
In contrast to the response behaviour in the ERiK-Surveys 2020, most of the questionnaires sent to youth welfare offices and family day-care workers were filled out in the first two months (see Figure 3.2-2). After the first week in the field, the participation rate for both surveys increased steadily over the course of January. After the start of the reminder process, about half of the completed questionnaires were answered by youth offices. In the case of family day-care workers, 42 % of questionnaires were handed in after this date. After the second reminder in 2022, 14 % of the youth welfare offices and 9 % of the family day-care workers filled out the questionnaire. In 2020, there was only one reminder round at a later point in time in the field. Around one third of the youth welfare offices and roughly half of the family day-care workers responded to the questionnaire after the start of this reminder round.

Most of the childcare providers also answered the questionnaires in the first two months after the initial contact

Figure 3.2-3 shows the field development for the ERiK-Survey 2022 of childcare providers. The survey institute sent the initial letters on 3 January 2022 by mail. The number of completed questionnaires rose steadily over the entire field period. By contrast, the participation rate in the ERiK-Surveys 2020 declined after the first two months of the field period.

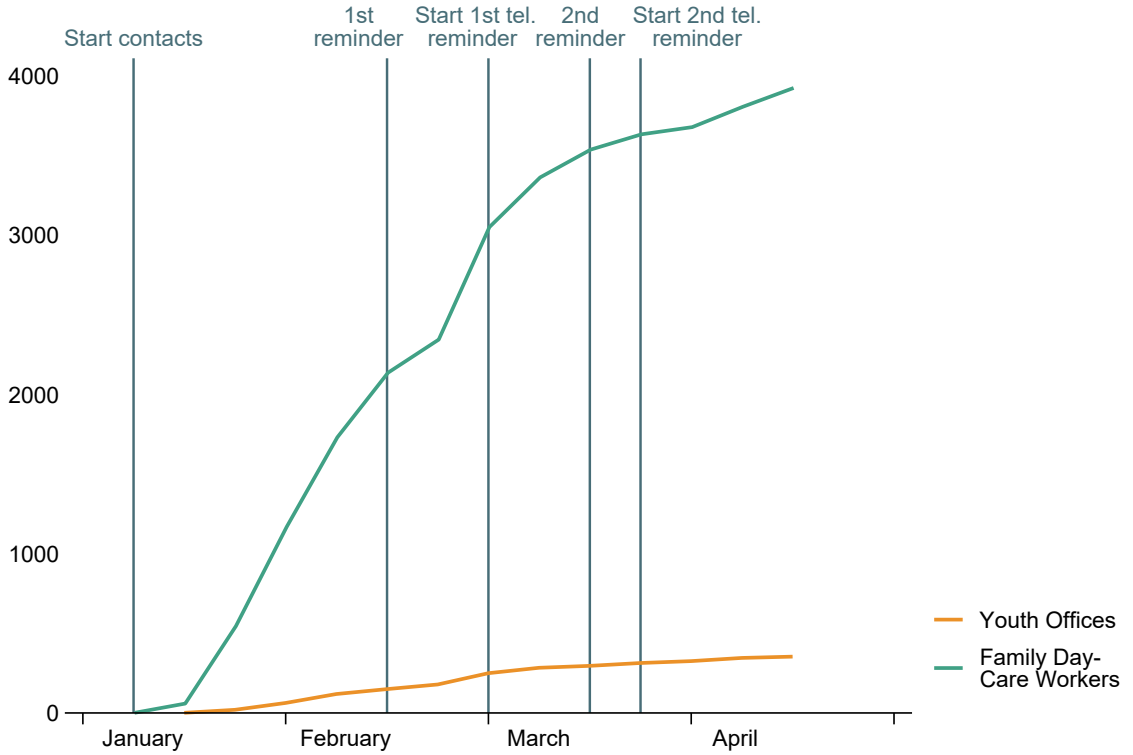
3 In three centres, seven children were interviewed, because the target child initially refused to participate and then expressed a desire to take part after a replacement child had already been interviewed.

Figure 3.2-1: **Weekly Number of (Partially) Completed Questionnaires: Directors and Pedagogical Staff**



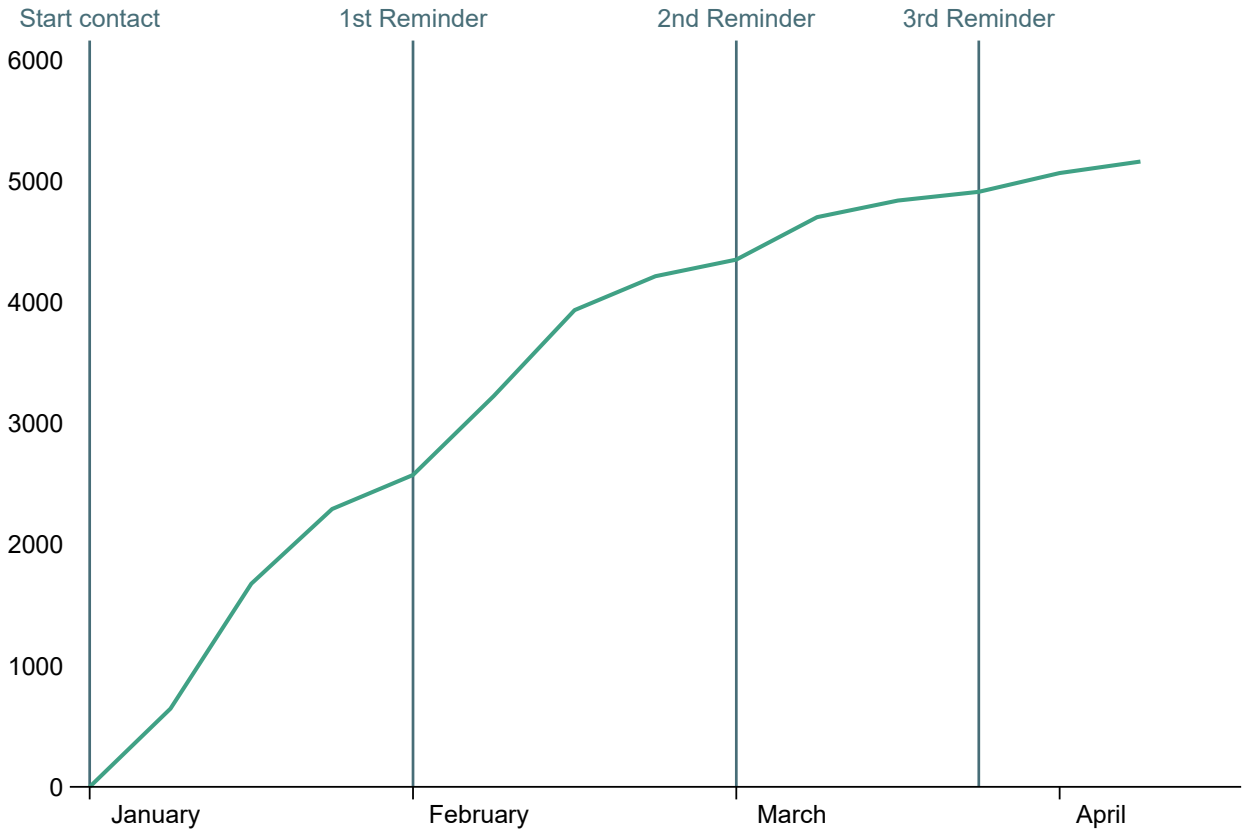
Source: DJI, ERiK-Surveys 2022: Survey of Directors, unweighted data, n=4832; DJI, ERiK-Surveys 2022: Survey of Pedagogical Staff, unweighted data, n=7116.

Figure 3.2-2: **Weekly Number of (Partially) Completed Questionnaires: Youth Welfare Offices and Family Day-Care Workers**



Note: Questionnaires for which no survey date is available are not shown (12 Youth Welfare Offices and 4 Day-Care Workers).
Source: DJI, ERiK-Surveys 2022: Survey of Youth Welfare Offices, unweighted data, n=417; DJI, ERiK-Surveys 2022: Survey of Family Day-Care Workers, unweighted data, n=5235.

Figure 3.2-3: **Weekly Number of (Partially) Completed Questionnaires: Childcare Providers**



Note: 4 questionnaires for which no survey date is available are not shown.
Source: DJI, ERiK-Surveys 2022: Survey of Providers of Childcare, unweighted data, n=6561.

Field period took place earlier than in the 2020 ERiK surveys.

The surveys of directors and pedagogical staff were conducted between February and April 2022, while those of family day-care workers, youth offices and providers were held between January and March 2022. Compared to 2020, the survey of directors and pedagogical staff took place earlier in the year in order to be closer in time to the survey of KJH statistics. The survey of parents (that was part of the children's survey) was conducted between March and April and the child survey between May and September 2022. More information on the field period of the separate parent survey KiBS can be found in the KiBS methodological report (Wieschke/Lippert/Kuger 2023).

All target groups were invited to participate by mail and/or e-mail. Where possible, respondents were also reminded of their participation by mail, telephone and/or e-mail. In 2022, the reminders were sent out at times closer to the survey date. For a detailed overview of all contact modes and reminder letters used, see Methodological Report III (Schacht et al. 2023).

Pedagogical staff, directors, and providers took part in the ERiK Surveys 2022 on average 26 to 30 days after initial contact, while family day-care workers and youth offices responded slightly later on average (32 to 40 days after initial contact). The parents participating in the child survey responded the fastest (approx. 20 days after initial contact). On average, the interviews with the children were conducted 110 days after the initial contact with the directors, if parental consent and some information about them could be obtained in the meantime. For the child interviews, one interviewer came to the day-care centre in person. For more information on the children survey, see von der Burg et al. (2021).

Table 3.2-1: **Field measures of the ERiK Surveys 2020 & 2022**

	Directors		Pedagogical Staff		Family Day-Care Workers		Youth Offices		Providers		Parents	Children
	2020	2022	2020	2022	2020	2022	2020	2022	2020	2022	2022	2022
Field period	April - August	February - April	April - August	February - April	May - September	January - May	May - September	January - March	May - September	January - March	March - April	May - September
Advertising/contact with umbrella organisations	x	x	x	x	x	x	x	x	x	x	-	-
Contact by post	x	x	x	x	x	x	x	x	x	x	x	x
Contact by e-mail	-	-	-	-	x	x	x	x	-		-	-
Contact by telephone	x	x	-	-	-	-	-	-	x	x	-	-
Incentives	x	x	x	x	x	x	-	-	x	x	x	x
Nonresponse Survey	x	x	-	-	-	-	-	-	x	x	-	-
Average no. of days between primary contact and participation (S.E.)	34.8 (0.43)	28.4 (0.28)	26.3 (0.22)	26.5 (0.22)	34.8 (0.60)	32.5 (0.32)	26.9 (1.26)	39.8 (0.97)	35.1 (0.51)	29.8 (0.27)	20.5 (0.28)	109.8 (1.09)

3.3 Number of Completed Questionnaires and Response Rates

In total, 4,674 directors, 7,019 pedagogical staff, 3,854 family day-care workers, 341 youth welfare offices, 4,710 providers and 479 children completed the questionnaires in full (see also in the synopsis 5.0-1). The number of completed questionnaires for providers and directors has increased significantly compared to the ERiK-Surveys 2020. For pedagogical staff, family day-care workers and youth offices, however, the number decreased somewhat (cf. Klinkhammer et al. 2022, p. 43).

The response rates (according to the American Association for Public Opinion Research (AAPOR) definition number 2, which includes complete and partial cases, for the German context: Stadtmüller et al. 2019) of the ERiK-Surveys 2022 (see Table 3.3-1) paint a mixed picture: There was a slight decrease in the relative response rate to 27 % (2020: 33 %) and 18 % (2020: 19 %) for directors and pedagogical staff. However, the absolute net number of responses from directors increased by approximately 900 cases and decreased by approximately 1,700 cases for pedagogical staff. For a detailed overview of the ERiK responses (including partial cases), see Table 3.3-1. The table also shows that response rates vary greatly across federal states (by up to 18 %): For directors between 19 % (Hamburg) and 37 % (Rhineland-Palatinate), for pedagogical staff between 9 % (Hamburg) and 27 % (Rhineland-Palatinate).[4]

For family day-care workers in 2022, the response rate halved to 9 % – with the number of net cases remaining the same – which is presumably mainly due to the significantly larger sample with almost exclusively online participation options[5] via double opt-in. This double opt-in method is a process where individuals confirm their subscription twice, first by signing up and then by verifying their intention through a confirmation message. This might be an obstacle for online participation (for the new design see Sections 5.2 and 6.2 in the ERiK method report III Schacht et al. 2023). The youth welfare offices must also forward the questionnaires to the family day-care workers. Presumably, some have not forwarded the questionnaires or have only forwarded them partially. Across the federal states, response rates vary between 4 % in Berlin and 16 % in Thuringia.

The response rate in the survey of youth offices fell sharply in 2022 (by almost 20 percentage points to 64 %), resulting in a reduction of 40 complete cases compared to 2020. There are major differences between the individual federal states. While the response rate in Mecklenburg-Western Pomerania is 38 %, it is 100 % in Hamburg and Saarland.

The high response rates for the child survey (between 42 % in Bavaria and 100 % in Saxony) can be attributed to the survey design. Only parents of day-care centres that had already participated in the directors survey in 2020 were asked for their agreement to participate in the additional child survey. Among parents who participated and gave their consent, which was only a small share of all parents, the participation of the associated children was correspondingly high at 69 %. This is particularly remarkable considering the long period of time that elapsed between the initial contact with the centres and completion of the child questionnaires. The response rates of the ERiK-Survey 2022 of children are shown in Figure 3.3-1 for better comprehensibility.

In contrast, the response rate for the ERiK-Survey of providers rose from 16 % to 24 %, with the number of cases therefore more than doubling. This could be due to an improved sampling frame[6], the promotion of the survey at specialist conferences or the reduction in the length of the questionnaire. Again response rates vary across the federal states. They range from 16 % in Hamburg to 32 % in Saarland (see Table 3.3-1).

At the same time, there are strong differences in the proportions of fully and partially completed questionnaires between the various ERiK target populations. Figure 3.3-2 shows these proportions, the frequency of the different types of (non)response and the eligibility to participate across the five regularly surveyed ERiK target groups. It can be seen that the proportion of unknown eligibility to participate has fallen in all

4 As explained in Section 5.2 and Table A.0-4 in the ERiK Methodological Report III (Schacht et al. 2023), the number of questionnaires for pedagogical staff per centre varied between one and six.

5 The response rate in online-only mode is significantly lower than when respondents have the choice between paper and online questionnaires (see Section 3.4).

6 The address lists for providers in 2020 were researched by a commercial provider. From 2022, the lists were obtained directly from the federal states.

Table 3.3-1: **Response Rates 2022 and 2020**

	Directors	Ped. Staff	FDW	Youth Offices	Providers	Children
2022						
Baden-Wuerttemberg	32 %	22 %	8 %	72 %	23 %	77 %
Bavaria	32 %	21 %	10 %	69 %	24 %	42 %
Berlin	22 %	13 %	4 %	58 %	22 %	78 %
Brandenburg	23 %	17 %	9 %	72 %	25 %	88 %
Bremen	24 %	14 %	13 %	50 %	22 %	73 %
Hamburg	19 %	9 %	11 %	100 %	16 %	59 %
Hesse	26 %	19 %	12 %	64 %	28 %	76 %
Mecklenburg-Western Pomerania	23 %	16 %	9 %	38 %	18 %	55 %
Lower Saxony	29 %	20 %	9 %	63 %	26 %	70 %
North Rhine-Westphalia	30 %	21 %	8 %	58 %	23 %	75 %
Rhineland-Palatinate	37 %	27 %	12 %	68 %	25 %	76 %
Saarland	25 %	17 %	13 %	100 %	32 %	55 %
Saxony	32 %	25 %	9 %	85 %	26 %	100 %
Saxony-Anhalt	25 %	16 %	11 %	79 %	28 %	81 %
Schleswig-Holstein	32 %	23 %	11 %	44 %	30 %	89 %
Thuringia	28 %	22 %	16 %	73 %	21 %	47 %
Total response rate	27 %	18 %	9 %	64 %	24 %	69 %
Total number of cases	4,832	7,116	3,927	366	5,166	490
2020						
Response rate	33 %	19 %	20 %	83 %	16 %	–
Number of cases	3,915	8,833	3,925	479	2,288	–

Note: Partial cases were counted here as having participated. Cases that are not eligible to participate were excluded from the calculation. Cases whose eligibility to participate is unknown were counted as eligible for the calculation. For the indirectly surveyed populations, the gross sample was reduced to those cases that had the opportunity to participate (to whom the survey was forwarded). FDW: Family day-care workers.

surveys compared to 2020.[7] There is also a relatively high proportion of partial or incomplete questionnaires among directors, providers and especially youth offices in 2020 and 2022. The large number of partial or incomplete surveys in the surveys of providers and youth offices can probably be explained by the fact that these cannot usually be completed by just one person, as several employees are often responsible for the different areas of responsibility surveyed in the questionnaires. At 0 % to 2 %, the proportion of cases not eligible to participate is at a similar level to 2020 for all target populations.

Table 3.3-2 shows how long respondents took to answer their respective questionnaires. When comparing the durations, it is apparent that the efforts to shorten the questionnaires for the ERiK-Surveys 2022 were effective in reducing the processing time for respondents.

Table 3.3-2: **Online Questionnaire Processing Time (in Minutes)**

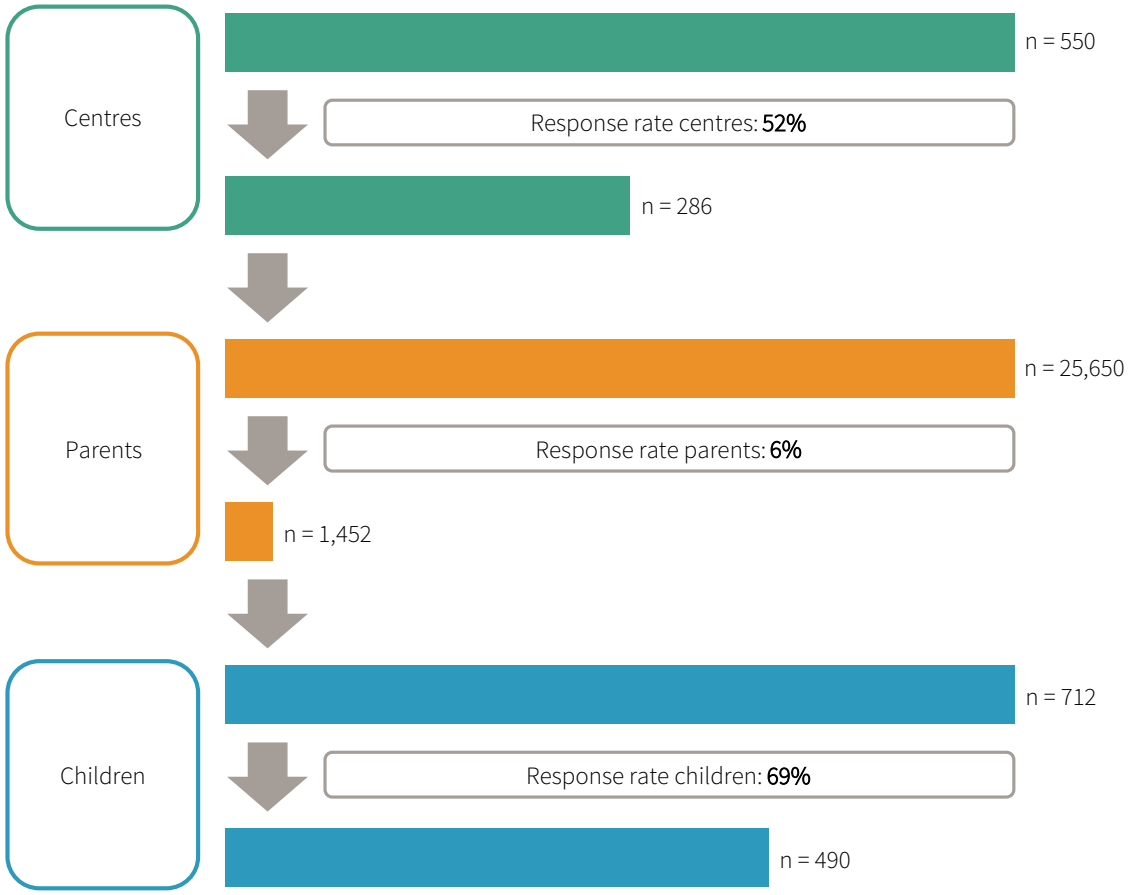
Population	2020		2022	
	Mean	Median	Mean	Median
Directors	55.5	52	46.5	43
Ped. Staff	34.4	31	31.5	28
Family Day-Care Workers	71.7	56	53.6	42
Youth Offices	114.3	81	62.5	40
Providers	64.5	47	47.8	32
Children	-	-	19.6	19

Note: Only complete cases were used to determine the processing time of the online questionnaires. Processing times exceeding 500 minutes were excluded from the calculations.

The reduction was moderate in the day-care centre surveys, with pedagogical staff taking about 3 minutes less and directors 9 minutes less to answer the questionnaire completely. Family day-care workers and providers of day-care centres were about 16 minutes faster in the ERiK-Surveys 2022 compared to 2020. For youth offices

7 In 2020 the percentages of unknown eligibility were 22 % for directors, 20 % for pedagogical staff, 7 % for family day-care workers, zero for youth offices and 83 % for providers.

Figure 3.3-1: **Response Rates for the ERiK-Survey 2022 of Children**



Note: The stated response rate of 6 % (which includes partial cases) among parents corresponds to the minimum response rate if it is assumed that all centres contacted (550) also forwarded the questionnaires to all parents. While it is very unlikely that this was implemented accordingly in all centres, the number of parent questionnaires forwarded cannot be determined. If only the participating centres (286) forwarded questionnaires to parents, the response rate for parents is more likely to be in the range of 12 %.

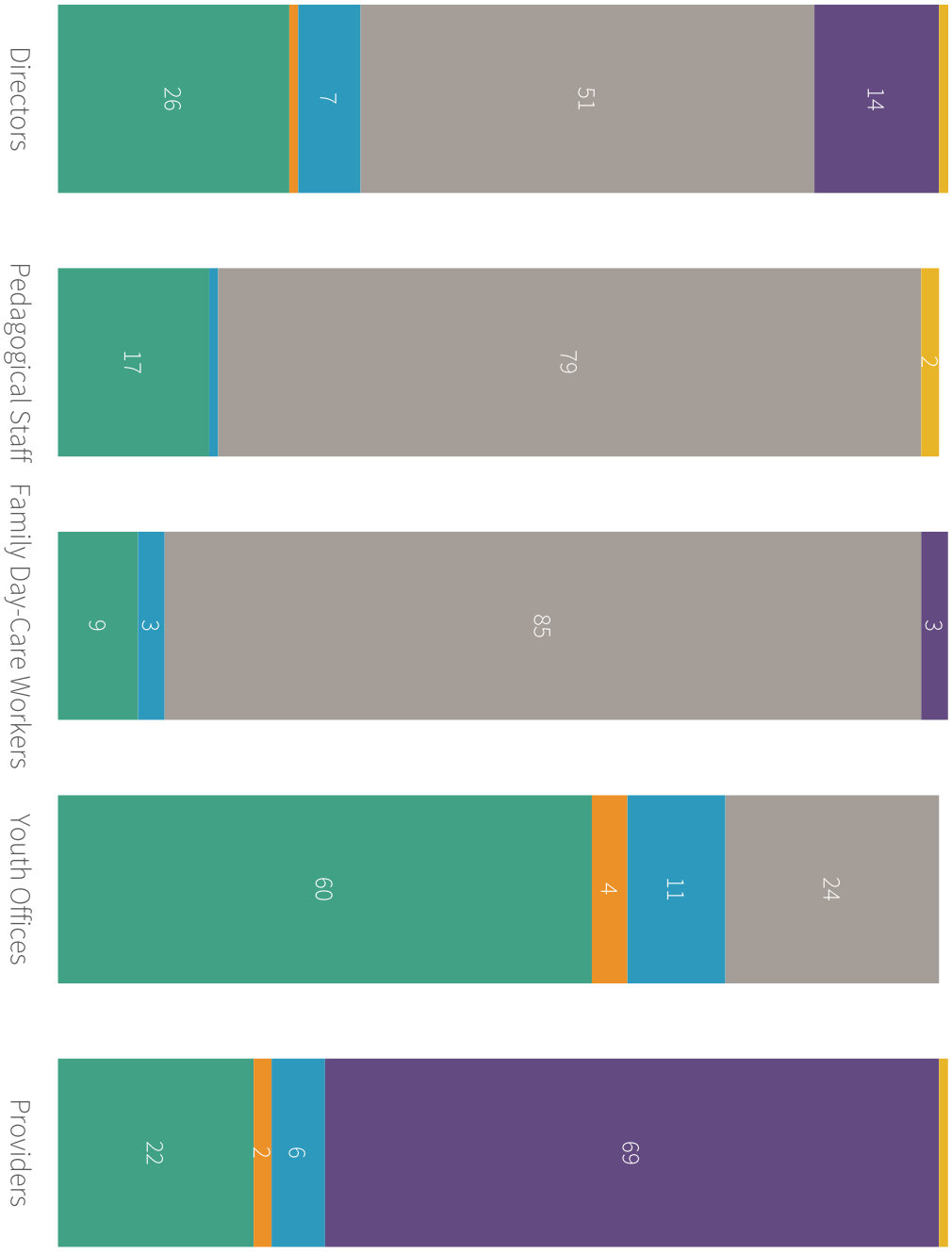
the reduction in answering time was substantial. Respondents completed the questionnaire in about half the time compared to the ERiK-Survey 2020 – in 62 minutes instead of 114 minutes.

3.4 Mode Choices

The response rates mentioned in the previous section (3.3) vary depending on population, the contact mode used and the options offered to respondents for participating (see Graph 3.4-1). In the ERiK-Surveys of individuals (directors, pedagogical staff, family day-care workers) the response rate was consistently higher in the concurrent mode where the sampled respondents received a paper questionnaire together with a link to the online questionnaire (P&O) instead of only a letter with the link (O).

For directors, the response rate was about 29 % for mode P&O and about 22 % for mode O. Among pedagogical staff, the difference is even more pronounced, with a 20 % response rate in the P&O group compared to just 9 % in the O group. Family day-care workers showed a similar trend: those who received only an email invitation (about 33,000 individuals) had a response rate of around 8 %, significantly lower than the 17 % response rate observed in the P&O group. For those who received a letter with a link to the online questionnaire, the response rate was intermediate, at approximately 12 %. See also 3.3 for further information. This indicates that at least in the ERIK-Surveys, sending out questionnaires on paper is one way to increase the response rates in the populations of directors, pedagogical staff and family day-care workers. There is no significant difference in response rates among providers, with around 24 % responding regardless of whether they received a paper questionnaire or just a letter with a link to the online survey.

Figure 3.3-2: **Shares of AAPOR Codes in the ERiK-Surveys 2022**



Note: Number of cases: Directors: 17,990; Pedagogical staff: 40,814; Family day-care workers: 43,759; Youth offices: 569; Providers: 21,625

Figure 3.4-1: **Response Rates by Invitation Mode in the ERiK-Surveys 2022**

Note: The response rates include complete and partial cases; ineligible cases are excluded. As all youth offices had the option to participate via paper or online questionnaire, they are not depicted in the graph.

Looking at the groups with concurrent mixed mode (P&O via mail), a similar pattern is apparent (see Graph 3.4-2). When given the choice between a paper and an online questionnaire, the majority of directors, pedagogical staff, and family day-care workers preferred the paper version. The percentage choosing the paper version ranged from 66 % for directors, over 75 % for family day-care workers and rising to 80 % for pedagogical staff. In the ERiK-Surveys of institutions, the mode choices were reversed: 68 % of youth offices and 72 % of providers chose the online questionnaire.[8]

3.5 Regional Coverage

Figures 3.5-1 and 3.5-2 show the regional coverage of the ERiK-Surveys 2022 in Germany. It should be noted, however, that the main focus of the ERiK-Surveys 2022 is not to be able to monitor differences at the district level[9]. The sample design[10] did not provide for comprehensive coverage of all districts, as the target population sizes would not be sufficient for this (for more information on sample sizes at regional level, see Section 4.4).

In four out of five surveys, participation in Germany was relatively evenly distributed across administrative districts

The colour gradations in Figure 3.5-1 represent the proportion of complete and partial questionnaires in each population according to official register data (Federal Statistical Office 2022) or, in the case of the provider survey, in accordance with the sampling frame at district level. The colour gradations are coded as follows: no participation is indicated by white,[11] the 10 % of the districts with the lowest participation rate are shown in the lightest colour. Further gradations are 25 %, 50 %, 75 % and 90 % of the districts. The top 10 % of the districts are thus represented by the darkest colour. These percentile cutoffs are used because the maximum proportion of the participating district population varies greatly across the different

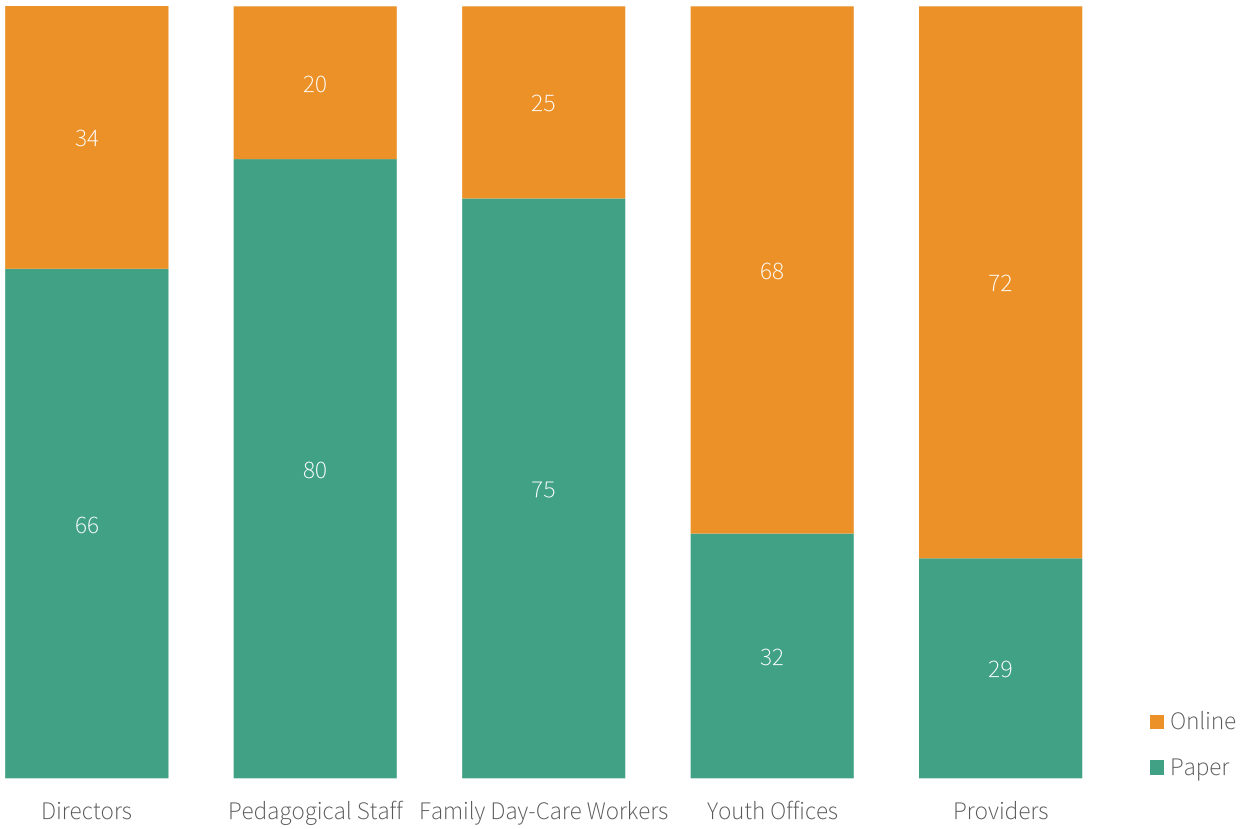
8 This is the case for complete or partial questionnaires. When looking at all questionnaires – including incomplete or aborted cases – the proportions for providers change slightly: in this case 59 % chose online and 41 % paper questionnaires. In the other surveys the differences are smaller (less than 5 %).

9 Germany is divided into 400 administrative districts. These districts consist of 294 rural districts (in German: 'Kreise' and 'Landkreise') and 106 urban districts (in German: 'Kreisfreie Städte' and 'Stadtkreise', the latter being towns or cities that constitute districts in their own right). This is one urban district less than in the ERiK-Surveys 2020 as the urban district of Eisenach was merged with the rural district Wartburgkreis in 2021 (European Commission 2021; Federal Statistical Office 2019).

10 Detailed discussions of the target population and sampling strategies of the ERiK-Surveys 2022 are provided in the ERiK Methodological Report III (Schacht et al. 2023).

11 The district is also marked white, if no person or institution was sampled in the respective district.

Figure 3.4-2: **Mode Choice in the concurrent mixed mode of the ERiK-Surveys 2022**



Note: The graph includes only cases that could choose between answering on paper or online and were either fully or partially completed; incomplete cases are excluded.

surveys (e.g. 75 % for providers vs. 0.7 % for pedagogical staff).

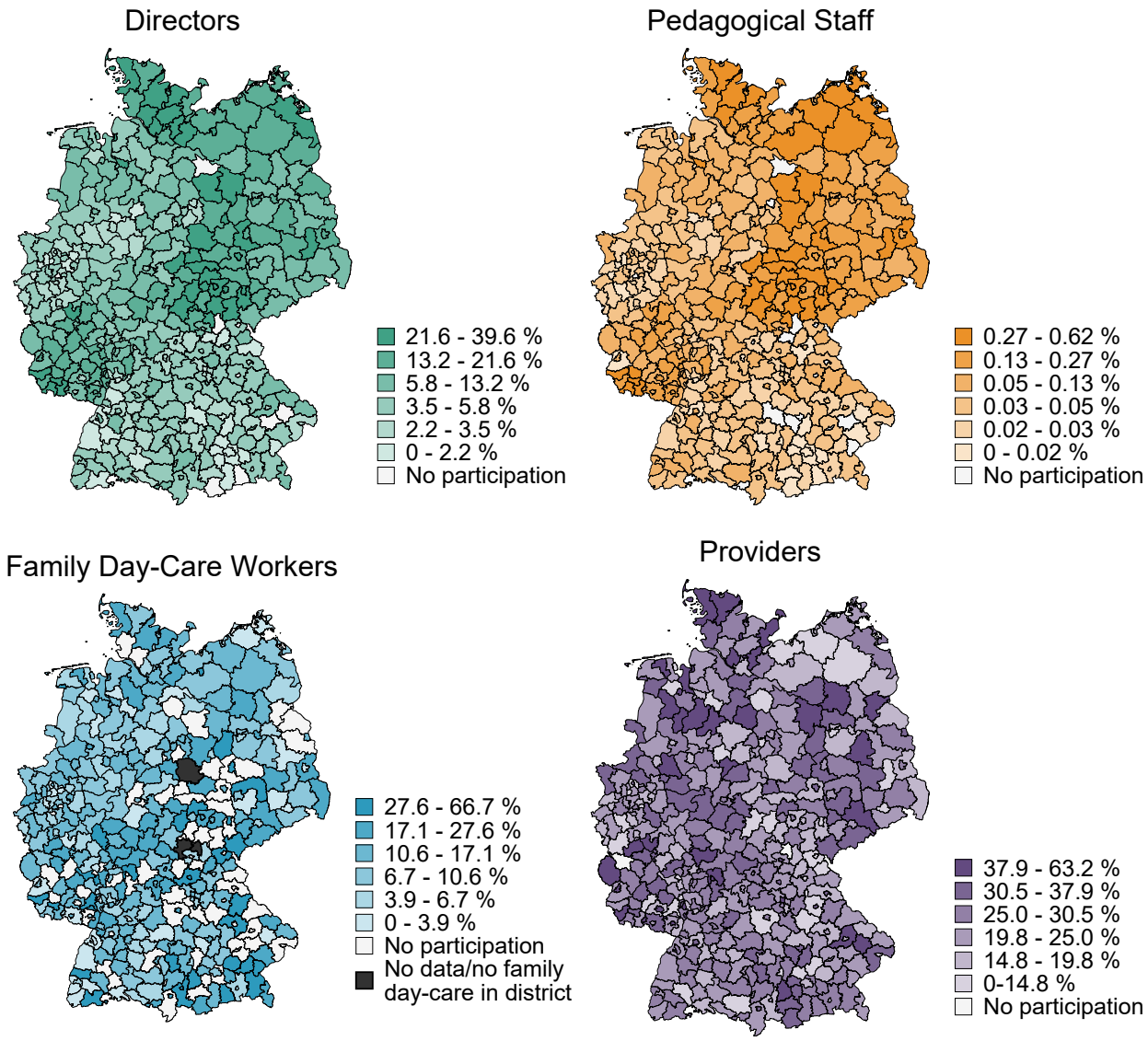
Directors and pedagogical staff of almost all districts and district-free towns took part in the respective survey. This is almost full coverage. In six districts and district-free towns, no directors or pedagogical staff took part. This represents an increase of two districts over 2020. In the northwestern and southern parts of Germany, the proportion of participants in the population is somewhat lower compared to the other regions. This pattern is very similar to the ERiK-Surveys 2020 (see Schacht et al. 2022).

In only two districts in Germany (Vulkaneifel in Rhineland-Palatinate and Sonneberg in Thuringia), did no providers take part in the survey. This represents a sizeable rediction compared with 2020, when 16 districts did not participate.

In 17,5 % of the districts and district-free towns in Germany, no family day-care workers participated. This figure is also considerably lower than in 2020 when 38 % of the districts and district-free towns failed to participate. This is most likely due to the fact that a random sample was taken in 2020 and only selected youth welfare offices were asked to forward questionnaires to the family day-care workers, whereas in 2022 a complete population survey was conducted among the family day-care workers.

The colour gradations in Figure 3.5-2 represent whether a youth welfare office participated (at least partially) in the survey at district level. In some districts and district-free towns, the youth welfare offices did not participate. As the map shows, there is no clear pattern to the regions in which youth welfare offices tended not to respond. The youth welfare offices that participated in the ERiK-Survey 2022 are more or less evenly distributed across Germany. In 29.5 % of the districts and district-free towns, no youth office responded to the survey. This is almost twice as many as in 2020, when participation stood at 15 %.

Figure 3.5-1: **Proportion of Questionnaires in each Population at District Level (in %)**



Note: Including partial and complete questionnaires; population sizes according to the KJH statistics 2023 (directors, ped. staff, family day-care workers) or the sampling frames (childcare providers) at district level. The boundaries of the categories are derived from percentiles of the distribution of the proportions. Rounded values are given in the legend for better readability.
Quelle: DJI, ERiK-Surveys 2022: Survey of Directors, n=4832; Survey of Pedagogical Staff, n=7116; Survey of Family Day-Care Workers, n=3927; Survey of Childcare Providers, n=5166; Research Data Centre of the Statistical Offices of the Federal States, Child and youth welfare statistics, 2023; unweighted data

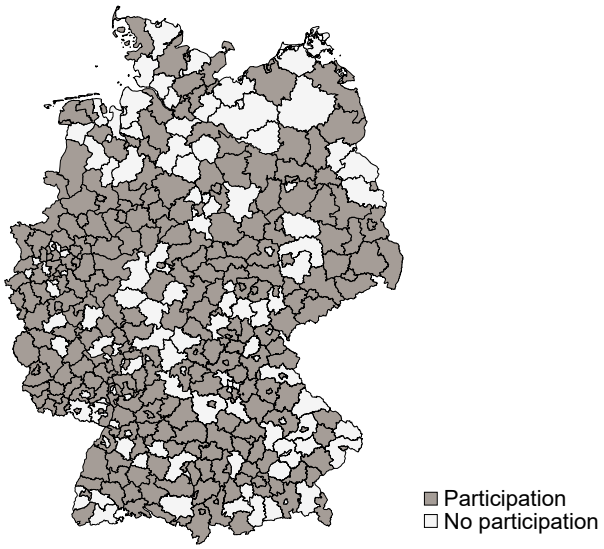
3.6 Nonresponse Surveys

In the course of two nonresponse surveys, directors and childcare providers who had not participated in the ERiK-Surveys 2022 were contacted by telephone shortly before the end of the fieldwork and asked about the reasons for their nonparticipation. The ERiK-Nonresponse-Survey of directors took place between 9 March and 14 April 2022, while the ERiK-Nonresponse-Survey of providers was carried out in the period from 4 to 22 April 2022 by the survey institutes.

Additional telephone contact with about 16,200 centres

A total of about 16,149 day-care centres were contacted during the telephone reminder process for directors. This is almost twice as many as in the ERiK-Surveys 2020, when 8,620 day-care centres were contacted. In about 4,372 centres (about 27 %), no director could be reached. In 847 of these cases, no valid telephone number was available. The day-care centres were contacted in total about 73,867 times, which implies

Figure 3.5-2: **Questionnaires of Youth Welfare Offices at District Level**



Note: Including partial and complete questionnaires
Quelle: DJI, ERiK-Surveys 2022: Survey of Youth Welfare Offices; n=366; unweighted data

an average of 4.6 contact attempts per day-care centre.

It was found that nearly 198 cases (about 1.2 %) were not part of the target population. These were day-care centres exclusively for school children or day-care centres that had been closed in the meantime. A total of 1,061 centres also indicated in the telephone reminder that they had already completed the questionnaire (6.6 %). A further 7,166 centres (about 44 %) indicated that they planned to participate in the near future. This is very similar to the ERiK-Survey 2020, where about 45 % stated that they were still planning to participate in the survey via online- or paper questionnaire (Schacht et al. 2022). In 4,372 (61 %) of these cases, i.e. those who indicated that they intended to participate via online questionnaire, an e-mail was sent with the web link to the questionnaire. This is considerably less than in 2020, when 81 % of the day-care centres that indicated that they planned to participate were sent an e-mail with the web link to the questionnaire.

In about 2,210 cases (approximately 14 %), participation in the main survey was refused. Short telephone interviews were conducted with 683 of these centres. These interviews for the ERiK Nonresponse-Survey of directors had an average duration of 3.3 minutes.

Telephone nonresponse interviews with about 683 directors lasted on average three minutes

After the telephone reminders of the directors were initiated, a further 1,895 questionnaires sent to the directors and 2,321 questionnaires to the pedagogical staff were completed. This means that about 39 % of the directors' questionnaires and about 33 % of the pedagogical staff questionnaires were answered after the telephone reminders started. For both the directors and the pedagogical staff, this is considerably more than in 2020, when 34 % of the directors' questionnaires and 7 % of the questionnaires sent to the pedagogical staff were answered after the telephone reminders started (Schacht et al. 2022).

Additional telephone contact with about 800 childcare providers

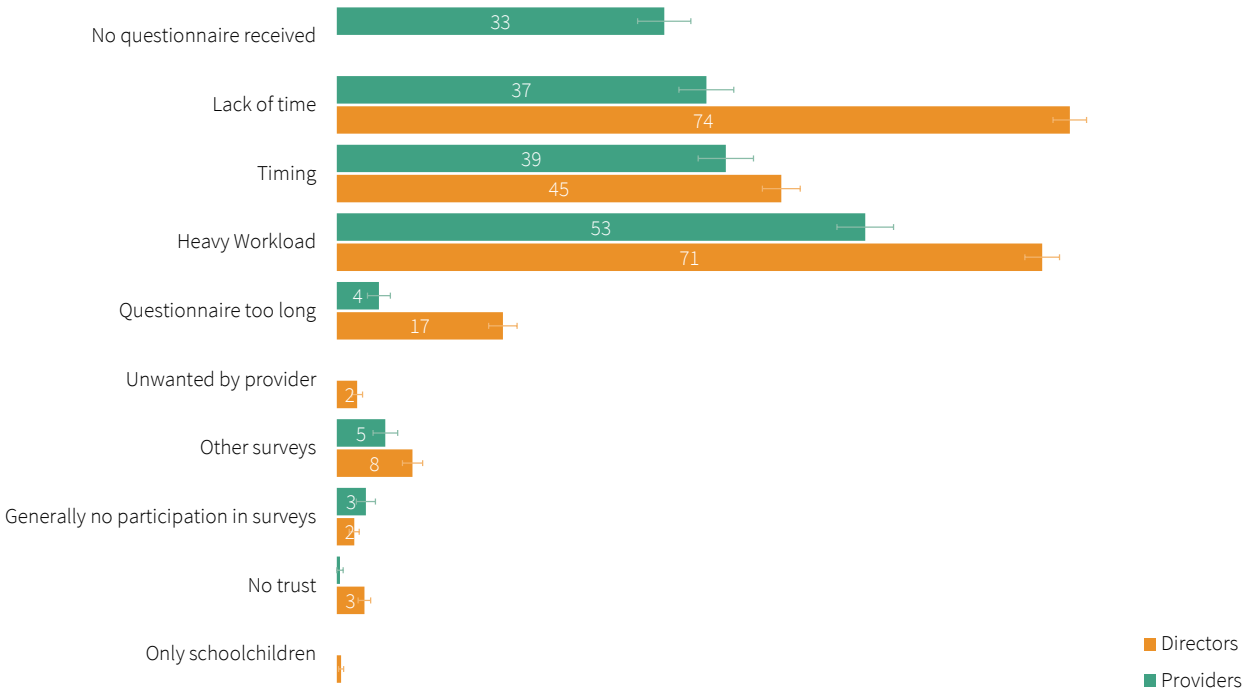
In total, 808 providers, for whom a telephone number was available, were contacted for the ERiK Nonresponse-Survey of providers, with an average of just over two contacts (min. one contact and max. eight contacts). The providers were randomly selected from the sample of providers who had not yet participated or actively refused to participate and for whom a telephone number was available.

Telephone nonresponse interviews with 307 providers lasted on average five minutes

In an interview lasting approximately five minutes, the providers were asked about structural characteristics (including the type and legal form of the provider and the number of day-care centres owned by the provider) as well as their reasons for non-participation. The interview was conducted with the person who received the letter of invitation including the questionnaire. Nonresponse interviews with 307 providers were conducted by telephone. In 2020, about 330 nonresponse interviews were conducted while 840 providers were contacted. No providers participated in the ERiK-Surveys 2022 after the telephone reminders were initiated.

The reasons for non-participation for both ERiK Surveys 2022 are shown in Figure 3.6-1. Time constraints were among the most important barriers to participation in the ERiK-Surveys 2022

Figure 3.6-1: **Results of the Two Nonresponse Surveys of Directors and Childcare Providers (in %)**



Notes: Up to three multiple answers possible; 8 directors and 27 providers that did not name any reason are not depicted. 'No questionnaire received' only asked for providers. 'Unwanted by provider' and 'Only schoolchildren' only asked for centres/directors.
Quelle: DJI, ERiK-Nonresponse-Surveys 2020: Directors, unweighted data, n=683; DJI, ERiK-Nonresponse-Surveys 2020: Providers, unweighted data, n=307

for both groups. For directors, a lack of time (74 %), heavy workload (71,%) or timing (45 %) were the main barriers to participation. For directors, these three reasons were also the most important ones in the ERiK-Surveys 2020. However, lack of time (in 2020: 69 %) and heavy workload (in 2020: 51 %) were mentioned less frequently in 2020. Also noteworthy is the fact that 17 % of the directors felt the questionnaire was too long. This is almost identical to 2020 (when it was 17 %, see Schacht et al. (2022)).

Compared to 2020, 'Lack of time' and 'Heavy workload' were mentioned more frequently as a reason for nonresponse

For childcare providers, a lack of time (37 %), timing (39 %) as well as a heavy workload (53 %) were prominent reasons for non-participation. In 2020, these figures were generally lower (lack of time: 29 %, timing: 25 %, heavy workload: 23 %). However, 33 % of the childcare providers said they did not receive a questionnaire, which is almost the same percentage as in 2020. The nonresponse survey also shows that in both population groups, it was rarely stated that they would not participate in surveys in general or that they would not trust the ERiK project. This has not changed in comparison with the ERiK-Surveys 2020 (Schacht et al. 2022).

More than one third of the providers in the ERiK-Nonresponse-Survey 2022 did not receive a questionnaire, and numerous day-care centers were contacted that do not belong to the target population. This indicates that the sampling frame needs to be improved for the ERiK-Surveys 2024. Note that in the ERiK-Surveys 2020, commercially acquired lists were used as a basis for contacting day-care centres and providers, while in the ERiK-Surveys 2022, the federal states were asked to supply address lists instead (see Schacht et al. 2023). It appears that the new sampling frame has not reduced the relative number of questionnaires that failed to reach the providers.

4. Data Quality of the ERiK-Surveys 2022

One goal of the ERiK study is to accurately determine the distribution and correlation measures for the ERiK target populations, aiming to evaluate the factors influencing the quality of Early Childhood Education and Care in Germany. To achieve this objective, it is crucial to consider potential sources of error at various stages of the survey process. An evaluation of some of these stages has already been conducted and documented in the ERiK Methodological Report III (Schacht et al. 2023). The chapter addresses the following key points:

1. The limited comparability of the survey years 2020 and 2022 due to changes in the target population for directors and pedagogical staff.
2. The selectivity of the surveys for family day-care workers and childcare centres according to official register data (the Child and Youth Welfare Statistics, in German: KJH – Kinder- und Jugendhilfestatistik), as well as childcare providers in general.
3. The weighting procedure is briefly explained and non-response weights are adressed in particular.
4. The samples are evaluated as to whether they are sufficiently large to make valid statements about subgroups in federal states.

Infobox 4.1 Representativeness

'Representativeness' is often regarded as a quality criterion for social science surveys. However, there is no clear mathematical-statistical definition of the term, but many ambiguous definitions (Kruskal/Mosteller 1980). Therefore, the term 'representativeness' is not used for the ERiK surveys. The data quality of the ERiK surveys 2022 is assessed on the basis of possible sources of error in the survey process, distributions and the sample size required for precise estimation of individual (point) estimators. Based on these considerations and with the addition of appropriate weightings, generalisable statements can be made about the framework conditions in child day-care in Germany on the basis of the ERiK data.

4.1 Limited Comparability of the 2020 and 2022 Surveys for Directors and Pedagogical Staff

As part of the ongoing development of the ERiK project, the target population for pedagogical staff was changed in 2022. Pedagogical staff with management responsibilities were excluded from the survey in 2022.[12] The decision to exclude pedagogical staff in managerial roles was based on the consideration that the experiences and viewpoints of this particular group might vary from those of non-managerial pedagogical staff. For detailed information on the target population in general and the specific changes, please see Schacht et al. (2023).

However, this decision resulted in limitations in the direct comparison of 2020 with 2022, as in the 2020 ERiK surveys, pedagogical staff with managerial tasks were invited to participate in the survey of pedagogical staff. In order to ensure comparability of the survey for the two years, the following procedure was applied: The pedagogical staff with managerial tasks were subsequently removed from the data of the survey on pedagogical staff in 2020 (corresponding to 16 % of net cases) in order to enable direct compari-

12 For a look at the specific modifications, the questionnaires for 2020 can be accessed at https://surveys.dji.de/?m=msw,0&sID=119, for 2022 at https://surveys.dji.de/?m=msw,0&sID=120. See also the screenshots of the letters to the directors A.0-2 and the questionnaire for pedagogical staff A.0-3 in the appendix.

sons with the target group of 2022.[13] Furthermore, the following adjustments to the design weights for 2020 were necessary in order to correct the selection probabilities for pedagogical staff:

$$DW_{PP_{new}} = \frac{N_{PP} - N_{PP-L}}{n_{PP}} \qquad (4.1)$$

where N_{PP} denotes the number of employees in a day-care centre, N_{PP-L} indicates the number of employees without management responsibilities and n_{PP} denotes the number of questionnaires in each day-care centre.

The calculation of the new design weight for 2020 largely corresponds to the calculation of the old design weight from 2020 (see Section 4.3). The difference is that pedagogical staff with management tasks are deducted from the total number of employees. The adjustment of the design weighting was only approximate, as this weight was derived from the variable *laj* (number of staff members in childcare facility and their occupational position) surveyed in the directors' questionnaire. It is possible that directors counted their pedagogical employees with management tasks as regular pedagogical staff. We can no longer identify those cases retrospectively. This problem arises not only in the design weighting, but for the entire population of pedagogical staff in 2020. This fact should be taken into account when analysing the 2020 dataset for pedagogical staff while excluding staff with management tasks.

Moreover, the revised definition of the target population of pedadogical staff is not compatible with the definition of pedadogical staff in the child and youth welfare statistics (KJH statistics), which means that no calibration to the marginal distributions of official data is possible (see also Section 4.3). Pedagogical staff with a small scope of management tasks or no formally defined management tasks cannot be identified in the KJH statistics. This means that persons with management tasks cannot be completely removed from the KJH figures and therefore there is no comparative group corresponding to the ERiK surveys. As a result, no statements can be made about the selectivity of this group in the ERiK Surveys 2022.

There were also minor changes to the target population in the directors survey. As in 2020, only managers with management as their primary responsibility were asked to complete the questionnaire. The definition of the directors target group for 2022 stated that the birthday method was applied when dealing with multiple managers possessing equal authority (management tandems) to guarantee randomized selection. This refinement of the target population can be considered as an advancement in the survey, yet it cannot be represented in this manner within the official KJH statistics. The KJH statistics indicate the number of management tandems, but the marginal distributions of auxiliary variables (e.g. scope of employment) for this subgroup are not known due to data protection.[14] However, the calibration is based on assumptions about the distribution of certain socio-demographic characteristics in the target population. If these assumptions are not accurate or complete, the weighted results may be biased (Bethlehem 2009). For this reason, additional calibration of the weights was also dispensed with in the ERiK directors survey.

The institution-level weights for the day-care centres remained unaffected by the target population changes and can be utilised for comparisons.

Although the results are comparable after the corresponding adjustments to the target population and weightings between 2020 and 2022, this fact should be taken into account when interpreting the results. One of the consequences of the adjustments is that results from 2020 presented in the current research report (Fackler et al. 2024) may differ from the results of earlier research reports.

Changes in the target populations of directors and pedagogical staff could lead to variations in the results compared to previous research reports.

As there is no suitable marginal distribution in the KJH data for pedagogical staff and directors, no additional calibration to the official statistics was carried out in these two surveys. Calibrat-

13 For classification of the aforementioned values: In the KJH statistics, the group of pedagogical staff with managerial tasks in the first and second area of work amounted to a total of 9 % for the whole of Germany in 2020; no information is available on staff with managerial tasks in the third area of work. The number of centres with non-contractually defined management tasks in 2020 amounted to an estimated national average of 8 % (Buchmann/Ziesmann/Drexl 2022); no further information is available on this group in the KJH statistics.

14 For clarification: According to the Child and Youth Services Statistics 2022, management tandems make up approx. 2.6 % of the total population or under 30 % of the population of directors throughout Germany.

ing the ERiK data to a marginal distribution that does not correspond to the respondents in the ERiK sample would introduce biases. Randomized sampling, a sufficiently large sample, and the correction for selection probabilities and non-response can provide unbiased estimators even without calibration (Bethlehem 2009). Nevertheless, the results may be biased with regard to the number of directors/pedagogical staff per federal state, educational degree and employment scope, as no calibration can be carried out without corresponding comparative data.

To illustrate the effects of the different types of weighting, the means for the variable 'Contractually agreed weekly working hours' per birth cohort are shown in A.0-4 in the appendix - unweighted and weighted for 2020 and 2022 for pedagogical staff and directors. We decided to use the variable 'Contractually agreed weekly working hours' as the sample was calibrated to this variable. Accordingly, the differences are likely to be greatest here. Nevertheless, the results differ only slightly: the point estimates hardly change and the variance is noticeably greater due to the weightings.

Infobox 4.2 Recommendations for Data Users

To summarise, we make the following recommendations to data users using the data for pedagogical staff:

- For comparisons between 2020 and 2022, the uncalibrated weights must be used (2020: nww_22; 2022: nww). The calibrated weights remain applicable for analyses exclusively focused on the year 2020 (nww_kal).
- For comparisons, the 2020 data must also be adjusted for pedagogical staff with management tasks (see variable *pul* in the updated dataset for pedagogical staff of the ERiK Surveys 2020).

Following recommendation applies for data users analysing the directors dataset:

- For comparisons between 2020 and 2022, the uncalibrated weights must also be used (2020: nww; 2022: nww). The calibrated weights remain applicable for analyses exclusively focused on the year 2020 (nww_kal).

The data for the ERiK Surveys 2022 and the data with the updated weights for 2020 will be made available at the DJI's Research Data Center in the course of 2024.

For a list of all available weights for all surveys, please refer to Appendix A.

4.2 Selectivity

In order to describe the selectivity in individual survey strands in ERiK, the marginal distributions of certain variables between the ERiK Surveys 2022 and the marginal distributions of official register data (KJH-Statistik Research Data Centre of the Statistical Offices of the Federal States 2022), or, in the case of the youth welfare office and provider surveys, the marginal distributions of the respective sampling frames, were compared[15]. In this way, any selections due to different sampling and participation probabilities were identified (for an overview and other sources of error see Groves 2004). All of the following marginal distributions differ at a statistically significant level ($p < .05$). Participants in the ERiK Surveys 2022 included:

1. fewer day-care centres with a lower number of approved places and more centres with many approved places (compared to the KJH statistics),
2. fewer family day-care workers with very small and more with very large numbers of children as well as more family day-care workers working in large day-care facilities with larger numbers of children (i. e. in German: Großtagespflege) compared to the KJH statistics, [16]
3. more public providers (in comparison to the sampling frame).

The marginal distributions of the youth welfare offices in the 2022 ERiK surveys did not differ significantly from the sampling frame in terms of the number of family day-care workers or the size of the municipality in which the youth welfare office is located. This was also not the case in the 2020 ERiK surveys.

15 There are also newer methods for providers (see Schacht/Ulrich 2024).

16 The official KJH statistics are only comparable with the ERiK sample for pedagogical staff and directors to a limited extent, as the target populations in both surveys were adjusted in 2022 (see Schacht et al. (2023) and 4.1). As a result, no comparison with the KJH statistics was made for these two target populations.

Similar selectivity among participants as in the 2020 ERiK surveys

Most of the differences found already existed in the ERiK-Surveys 2020 (Schacht et al. 2022). However, there were also differences between the population and the sample that existed in 2020 and no longer exist in 2022: In contrast to the ERiK Surveys 2020, no statistically significant difference was found with regard to the type of provider (private versus public providers) of the centres compared to the marginal distributions of the KJH data, i.e. one can speak of an improvement in the data basis in the survey of providers with regard to this characteristic. One of the reasons for this is that the quality of the address lists has presumably improved compared to 2020 due to the use of address lists from the German federal states (see Schacht et al. 2023).

Differences between the marginal distributions of the official KJH statistics and the ERiK sample are merely an indication that systematic differences exist between the real target population and the sample. It is important to note that the differences in the marginal distributions alone are not sufficient to draw concrete conclusions about the significantly more complex real multivariate distributions. These differences are accounted for in ERiK by extensive weighting procedures, that include selection probabilities (see Section 4.3).

4.3 Weighting

To generalise findings from the ERiK Surveys 2022, we adjust for variations in sampling and participation probabilities using a weighting method. This involves combining design weights to account for sampling complexities, nonresponse weights to address differences in survey participation rates, and calibrated weights based on population distributions. For interested readers, we recommend the detailed literature, e. g. Valliant/Dever/Kreuter (2013).

Design Weights

The complex sampling designs (see Section 5 in Schacht et al. 2023) for day-care centres, their pedagogical staff, children and family day-care workers are adjusted by using design weights.

The design weighting is carried out at the level of the centres, directors and pedagogical staff. Based on the centres, the selection probability is calculated as the quotient of the total number of centres per federal state N_E and the number of centres in the sample n_E.

$$P(E) = \frac{N_E}{n_E} \tag{4.2}$$

For the selection probability of the directors, a quotient is formed from the total number of directors per centre N_L and the number of directors who are to participate n_L (usually corresponds to 1, unless there are several directors, who share responsibilities equally, working in one centre) and multiplied by the centre weight.

$$P(L) = P(E) \times \frac{N_L}{n_L} \tag{4.3}$$

The selection probability for pedagogical staff is calculated analogously and a quotient is formed from the total number of pedagogical staff per centre N_P and the number of pedagogical staff who are to participate n_P, which is multiplied by the centre design weight:

$$P(P) = P(E) \times \frac{N_P}{n_P} \tag{4.4}$$

As the surveys of providers, youth welfare offices and family day-care workers are designed as complete population surveys (Vollerhebung), no design weighting is applied here.[17]

Nonresponse Weights

In order to make valid population inferences based on the ERiK-Surveys 2022, the samples are drawn on the assumption that all individuals in the different target populations have the same non-null inclusion probability within their respective population. If this is the case, the samples drawn are referred to as probabilistic. In practice, however, and also in our case, such an assumption is not realistic. In the ERiK Surveys 2022, the samples are drawn from volunteer participants that may have different reasons to participate or not to participate (Ferri-García del Mar 2020). Note that 'participation' in the youth office and provider survey means that the survey has been fully completed. In the directors and family day-care workers survey, however, 'participation' means full and partial completion of

17 The non-response weight for family day-care workers in 2022 corresponds to the design weight for family day-care workers in 2020. Only the designation has been changed here.

the questionnaire. This distinction was made in order to keep the samples balanced and thus improve model performance. As participants are usually different from the non-participants, a selection bias must be assumed. To correct for this bias, propensity score adjustment is performed through inverse propensity weighted estimators. For this, a single model is formulated for each population to estimate response probabilities. The weight adjustments are performed on every outcome variable of interest. However, it is recommended that the models to estimate the response probabilities should be different for different outcome variables (Ferri-García et al. 2022). In surveys with many outcome variables, such as the ERiK-Surveys 2022, this is not feasible. Thus, a model has to be fitted that includes all covariates that could be of importance for any outcome variable. In contrast to logistic regression models, it is possible to include many covariates with complex structures in a propensity score model if non-parametric machine learning methods are used (Buskirk/Kolenikov 2015; Lee/Lessler/Stuart 2010; Watkins et al. 2013; Westreich/Lessler/Funk 2010). A logistic regression model, on the other hand, is not very robust to model misspecification. This means that biased probability estimates may be produced if the linearity assumptions are violated or relevant interactions are not included (Gelein/Haziza/Causeur 2018). This makes logistic regression models for propensity score estimation susceptible to underfitting (see Hastie/Tibshirani/Friedman 2009).

We therefore use Random Forests (Breiman 2001) for each population of the ERiK-Surveys 2022 to estimate response probability. A Random Forest is an ensemble learning method that operates by constructing a multitude of decision trees and outputting the mode of the classes (here: participation or non-participation) within the terminal nodes of the trees for every individual in the sample. The covariates are used for splitting the sample to maximally reduce variance with respect to the outcome variable (here: participation) within the terminal nodes of a tree. A Random Forest introduces randomness between the trees in two ways: by selecting a random subset of covariates for splitting nodes and by using bootstrap samples of the data for building each tree (Kern/Klausch/Kreuter 2019). However, including many covariates in a non-parametric algorithmic model such as a Random Forest bears the risk of overestimating the variance of response probabilities, as it may be easier to overfit the model if it becomes too complex (Ferri-García et al. 2022). On the other hand, underfitting a propensity model may lead to underestimation of the variance of response probabilities.

The covariates used in the Random Forest models are described in row 2 of Table 4.3-4. Using Random Forests makes it possible to include a great number of covariates in the model. We therefore merged the sampling frames of the respective populations with administrative data on either the district level (Bertelsmann Stiftung 2024) or municipality (Bundesinstitut für Bau-, Stadt- und Raumforschung 2022) level. Overall, 156 covariates (in machine learning terms: *features*) are used for the non-response models for day-care centres, pedagogical staff, and providers. For youth welfare offices and family day-care workers, 149 features were used for each population. Note that there was no sampling frame for family day-care workers available to us. Thus, we predicted the probability of a youth office forwarding the questionnaire to the family day-care workers in their district. The inverse of these probabilities were then used as non-response weights for family day-care workers, as in the ERiK-Survey 2020 (Schacht et al. 2022).

The random forest algorithm requires users to configure several hyperparameters. This means that Random Forests can differ with respect to, for example, the total number of trees that the forest consists of, the number of observations randomly selected for each tree, or the number of variables randomly chosen for each split. To optimise the Random Forests, hyperparameter tuning is performed for each population by conducting a random search on various parameter settings with threefold cross-validation (see Probst/Wright/Boulesteix 2019). We defined a grid with 6,480 hyperparameter combinations shown in Table 4.3-1. We computed 200 iterations per population. This means that 3 × 200 Random Forests were fitted with random hyperparameter combinations for each of the sampling frames of the ERiK-Surveys 2022. Each model was fitted to the training fold of the data (with threefold cross-validation this is two thirds of the sample), each fitted model was used to predict response probabilities for the test fold (one third of the sample).

The mean of the ROC-AUC values of the three folds depicts the generalisation error of a Random Forest application. Out of the 200 iterations, the Random Forest hyperparameters with the best generalisation error are chosen. These hyperparameters are shown in Table 4.3-2. Lastly, the Random Forest with the best hyperparameters is fitted to the whole dataset and used to predict response probabilities for the same dataset. The inverse of these probability estimates are then used as non-response weights.

Table 4.3-1: **Random Forest Hyperparameters**

Name	Description	Values
n_estimators	number of trees in the forest	[200, 400, 600, 800, 1000, 1200, 1400, 1600, 1800, 2000, 2200, 2400, 2600, 2800, 3000]
min_samples_split	minimum number of samples required to split an internal node	[2, 5, 10]
min_samples_leaf	minimum number of samples required to be at a leaf node	[1, 2, 4]
max_features	number of features to consider when looking for the best split	['sqrt', None]
max_depth	maximum depth of the tree	[10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, None]
bootstrap	Whether bootstrap samples are used when building trees	[True, False]

Notes: The Names of Hyperparameters in the first column are the same as the attributes for the RandomForestClassifier function in the Python package scikit-learn, see www.scikit-learn.org

To evaluate the quality of the response probability estimates of the ERiK-Surveys 2022, we compare them to probability estimates from logistic regression models. For this, we recreate the logistic regression models used for the ERiK-Surveys 2020. We compare the estimated response probabilities (or forwarding probabilities for family day-care workers) predicted by the Random Forests with the probabilities modelled by the logistic regression models. The results are shown in Table 4.3-3. With respect to the ROC-AUC, the Random Forests clearly outperform the logistic regression models. Due to the small youth office sampling frame, the ROC-AUC values for the Random Forests for youth offices and family day-care workers are very high although the hyperparameters of the model are chosen to not overfit the data.

Calibrated Weights

In the final stage, the combined design and non-response weights underwent calibration using established population distributions. This process aimed to reduce variance and ideally address any remaining nonresponse or coverage errors not previously corrected. The ERiK project utilised population distributions from the official KJH statistics of 2022 and implemented iterative proportional fitting, also known as raking, as the calibration method (Deming/Stephan 1940). This approach is typically employed when only marginal distributions of the population are available for the calibration variables, rendering poststratification unfeasible (Sand/Kunz 2020).

The non-response weights of the following populations were calibrated using selected distributions of the KJH statistics 2022 (Research Data Centre of the Statistical Offices of the Federal States 2022) using 'iterative proportional fitting' (IPF, Zaloznik 2011) to the following characteristics:

- Family day-care workers (FDW): number of FDW per federal state, number of children in FDW care, qualification of FDW
- Youth welfare offices: Number of youth welfare offices per federal state, number of inhabitants, number of FDW, number of centres (in each case per youth welfare office district)
- Provider: At centre level: number of centres per federal state, provider type. At provider level (calibration to sample list[18]): Number of providers per federal state, provider type, provider size
- Children: number of children in care per federal state, age, gender, care time, migration background

18 Unfortunately, no complete address lists or official KJH figures for the year 2022 are available for providers. As a result, the DJI has requested the addresses of the providers from the responsible state administrations. Calibration was carried out using the information in these address lists. No guarantee can therefore be given for the completeness of these lists.

Table 4.3-2: **Selected Hyperparameters of Random Forests**

	n_estimator	min_samples_split	min_samples_leaf	max_features	max_depth	bootstrap
Day-Care Centres	2600	10	4	sqrt	10–index5	True
Pedagogical Staff	1200	2	4	None	10	True
Family Day-Care Workers	800	2	2	sqrt	20	False
Youth Offices	400	2	4	None	20	True
Providers	1400	2	4	sqrt	10	True

Notes: The Names of Hyperparameters in the header are the same as the attributes for the RandomForestClassifier function in the Python package scikit-learn, see www.scikit-learn.org

Table 4.3-3: **Results of Random Forest vs. Logistic Regression**

	target pop. size	Random Forest			Logistic Regression		
		no. of feat.	ROC-AUC	variance	no. of feat.	ROC-AUC	variance
Day-Care Centres	17991	156	0.74	0.005	18	0.58	0.002
Pedagogical Staff	40824	156	0.76	0.005	18	0.6	0.002
Family Day-Care Workers	574	149	1	0.015	4	0.5	0.003
Youth Offices	574	149	0.99	0.06	4	0.55	0.003
Providers	21426	156	0.78	0.003	23	0.56	0.001

As already mentioned, the weights for directors and pedagogical staff are not calibrated, as the definitions of the target populations do not match the KJH statistics (see Section 4.1). All final weights are trimmed once at the 99 % percentile. The final weight usually refers to the calibrated trimmed weight; if calibration has not taken place, the non-response weight is trimmed and considered to be the final weight. The final weights are available to users both trimmed and untrimmed. Normalisation of the weights is always population-based, i.e. based on the size of the total population determined according to the KJH statistics or the sampling frame. Normalisation adjusts the weights to ensure that they sum to the population size. This helps to correct for any discrepancies between the sample and the population, ensuring that estimates derived from the weighted sample are reliable (Sand/Kunz 2020). Due to the changes to the target population definitions of directors and pedagogical staff, some adjustments to the 2020 weighting factors were necessary as described in Section 4.1. These changes to the weights also allow for comparisons with the 2020 data using the new 2022 definition of the target population, thereby ensuring comparability across the years.[19] For an overview of the calculation of the weights from the survey year 2022, please see Table 4.3-4. For an overview of all weights provided in all datasets, please see Appendix section A.

19 These weighting factors are expected to be made available in a new version of the 2020 data in the research data centre (Forschungsdatenzentrum) of the DJI in 2024 as part of the publication of the ERiK-Surveys 2022 data.

Table 4.3-4: **Overview of the Weighting Process for the ERiK-Surveys 2022**

	Centres / Directors	Pedagogical Staff	Family Day-Care Workers	Youth Welfare Offices	Providers
Design weight (DW)	total number of centres in a federal state divided by the number of centres contacted in the federal state	multiplication of 1. the design weight of centres by 2. the inverse proportion of pedagogical staff contacted of the number of existing pedagogical staff in the centre	none (complete survey)	none (complete survey)	none (complete survey)
Nonresponse weight (NW)	Random forest with 1. federal state, 2. municipality size, 3. type of provider (8 categories), 4. provider size, and a total of 57 regional variables at municipality level and 47 regional variables at district level	Random forest with 1. federal state, 2. municipality size, 3. type of provider (8 categories), 4. provider size, and a total of 57 regional variables at municipality level and 47 regional variables at district level	Youth welfare office forwarding probability: Random forest with 1. federal state, 2. municipality size, 3. completeness of the youth welfare office survey, 4. participation in the previous survey year, and a total of 57 regional variables at municipality level and 47 regional variables at district level	Random forest with 1. federal state, 2. municipality size, 3. completeness of the survey, 4. participation in the previous survey year, and a total of 57 regional variables at municipality level and 47 regional variables at district level	Random forest with 1. federal state, 2. municipality size, 3. type of provider (8 categories), 4. provider size, and a total of 57 regional variables at municipality level and 47 regional variables at district level
Calibrated weight (CW)	for centres: 1. number of centres per type of provider (3 cat.) per federal state, 2. number of places for children (8 cat.). For directors: no calibration weight because of target population change	no calibration weight due to target population change	1. number of FDW per federal state, 2. number of children in family day-care (6 cat.), 3. highest degree and hours of qualification course of FDW (8 cat.)	1. number of YWO per federal state, 2. number of inhabitants per YWO district (5 cat.), 3. FDW per YWO district (5 cat.), 4. day-care centres per YWO-district (5 cat.)	for centres: 1. number of centres per federal state, 2. type of provider (7 cat.). For providers: 1. number of providers per federal state, 2. type of provider (7 cat.), 3. size of provider (4 cat.)
Trimming of final weights	centres: < 1 and > 99 % directors: > 99 %	> 99 %	> 99 % at NW and final weights	> 99 %	> 99 %
Names of final weighting variables	centres: nwe | directors: nww	nww	nww	nww	centres: nwe | providers: nww

Notes: YWO: Youth welfare office; FDW: Family day-care worker. The non-response weight for FDW in 2022 corresponds to the design weight for FDW in 2020. Only the designation has been changed here. The full list of regional variables used in the nonresponse weighting can be found in Appendix: A.

4.4 Informative Value at the Federal and State Levels

The ERiK Surveys 2022 have no limitations for the analysis at federal level for the whole of Germany, i.e. generalisable conclusions for the target groups can be drawn from the weighted ERiK data for the whole of Germany.

No restrictions on the informative value for the whole of Germany

However, evaluations for individual subgroups or state-specific evaluations (Bundesländer) may have limitations, as it is often difficult to achieve the required sample sizes for small target groups or regionalised analyses. Small samples are associated with a larger average sampling error (Groves/Lyberg 2010). This can impair the statistical reliability and generalisability of the results at federal state level. For the ERiK Surveys 2022, calculations on the required sample size for selected characteristics of all 16 federal states were carried out in advance, taking into account different uncertainty and error parameters. The following equation is used to estimate the sample size for a proportion of a population:

$$n \geq \frac{P(1-P)}{\frac{e^2}{z^2}\frac{\alpha}{2} + \frac{P(1-P)}{N}} \tag{4.5}$$

where n denotes the required sample size, N is the population size, e is the margin of error, which is expressed as a percentage and indicates how closely the responses from the chosen sample will approximate the true value of the overall population, z is the z-score corresponding to the desired confidence level and p represents the average deviation of individual values within a sample from the sample mean. Without any prior knowledge, the value p is usually set to 0.5 (Kauermann/Küchenhoff 2010). Depending on the desired precision and prior knowledge, different sample sizes can be expected. For the ERiK Surveys 2022, conservative estimates were generally chosen, necessitating relatively large sample sizes. For a detailed description of the sampling design procedure for all five ERiK Surveys 2022 and the estimated sample sizes, see Section *5. Sampling Designs* in Schacht et al. (2023). For the actual number of cases realised, see Table 3.3-1 in Section 3.3.

Nevertheless, small sample sizes are unavoidable in some federal states, as there are often only a few centres or institutions available for the survey, especially in small federal states (for a detailed discussion on this topic, see Section 4.3 in Schacht et al. 2022).

Moreover, the clustering in the surveys for pedagogical staff and family day-care workers should be taken into account when analysing the data. Cluster effects refer to the tendency of observations within the same cluster to be more similar to each other than to observations in other clusters (Kauermann/Küchenhoff 2010). In the ERiK Surveys the pedagogical staff are clustered in child day-care centres and family day-care workers are clustered in youth welfare offices. This clustering leads to a potential underestimation of variability in both surveys. Traditional statistical methods often assume independence of observations, but when data is clustered, this assumption is violated. As a result, standard errors may be underestimated, leading to overly narrow confidence intervals and inflating the likelihood of finding statistically significant results. For a detailed discussion on how the intraclass-correlation can affect the estimates on the ERiK Surveys, please see Section *4.3 Informative Value at the Federal and State Levels* in Schacht et al. (2022). In these two survey strands, data users are advised to take into account the clustering and the survey design by using corresponding software packages.

Infobox 4.3 Recommendations for Data Users

In the surveys of pedagogical staff and family day-care workers, data users are advised to account for the clustering and survey design by using appropriate software packages (e.g. the *Survey*package in R (Lumley 2023) or the *survey* functions in Stata).

Following consultation with experts from the field of ECEC and statistics, a number of rules for dealing with low case numbers were developed for the ERiK Surveys 2020. On the one hand, these are based on considerations regarding the accuracy and informative value of the estimates for the particular survey, the fulfilment of (normal) distribution assumptions and, on the other hand, the preservation of the data anonymity. For an

overview of the regulations in 2020 and a detailed discussion of the topic, see Schacht et al. (2022). The regulations have been slightly simplified in 2022 compared to 2020 to make them easier to understand and are now summarised as in Table 4.4-1.

Figure 4.4-1 contains an overview of the informative value of the ERiK Surveys of all federal states for the years 2020 and 2022 across all surveys. The change in the target population of pedagogical staff (see Section 4.1) had no retrospective effect on the evaluability of individual federal states in 2020.

Table 4.4-1: **Rules for dealing with small case numbers**

Survey	Major limitations	Minor limitations
Directors	($n \leq 75$)	($n/N < 2\,\%$)
Pedagogical Staff	($n \leq 75$)	($n/N < 0.5\,\%$)
Family Day-Care Workers	($n \leq 50$) oder ($n/N < 5\,\%$)	($n/N < 10\,\%$)
Youth Offices	($n < 10$) oder ($n/N \leq 50\,\%$)	–
Providers	($n < 50$) oder ($n/N < 10\,\%$)	–

Note: n: Sample size; N: Population size; Calculations only based on complete cases

In 2022, there were no restrictions for the surveys among directors. However, for pedagogical staff, the response rate in Baden-Württemberg, Bavaria, and North Rhine-Westphalia was relatively low (less than 0.5%), considering the population size in those regions. Nevertheless, the number of cases was sufficient, and analyses can be conducted without limitations, though the low response rate should be taken into account.

A similar situation occurred for family day-care workers in the states of Baden-Württemberg, Mecklenburg-Western Pomerania, Lower Saxony, North Rhine-Westphalia, and Saxony. In these regions, the response rate was very low, measuring less than 10 % relative to the population size in this group. In Berlin, Hamburg, Saarland, Saxony-Anhalt, and Thuringia, the response rate among family day-care workers was below 5 % of the population size, or fewer than 50 family day-care workers participated, making it advisable to use caution when interpreting the data from these states, as the results might not be reliable.

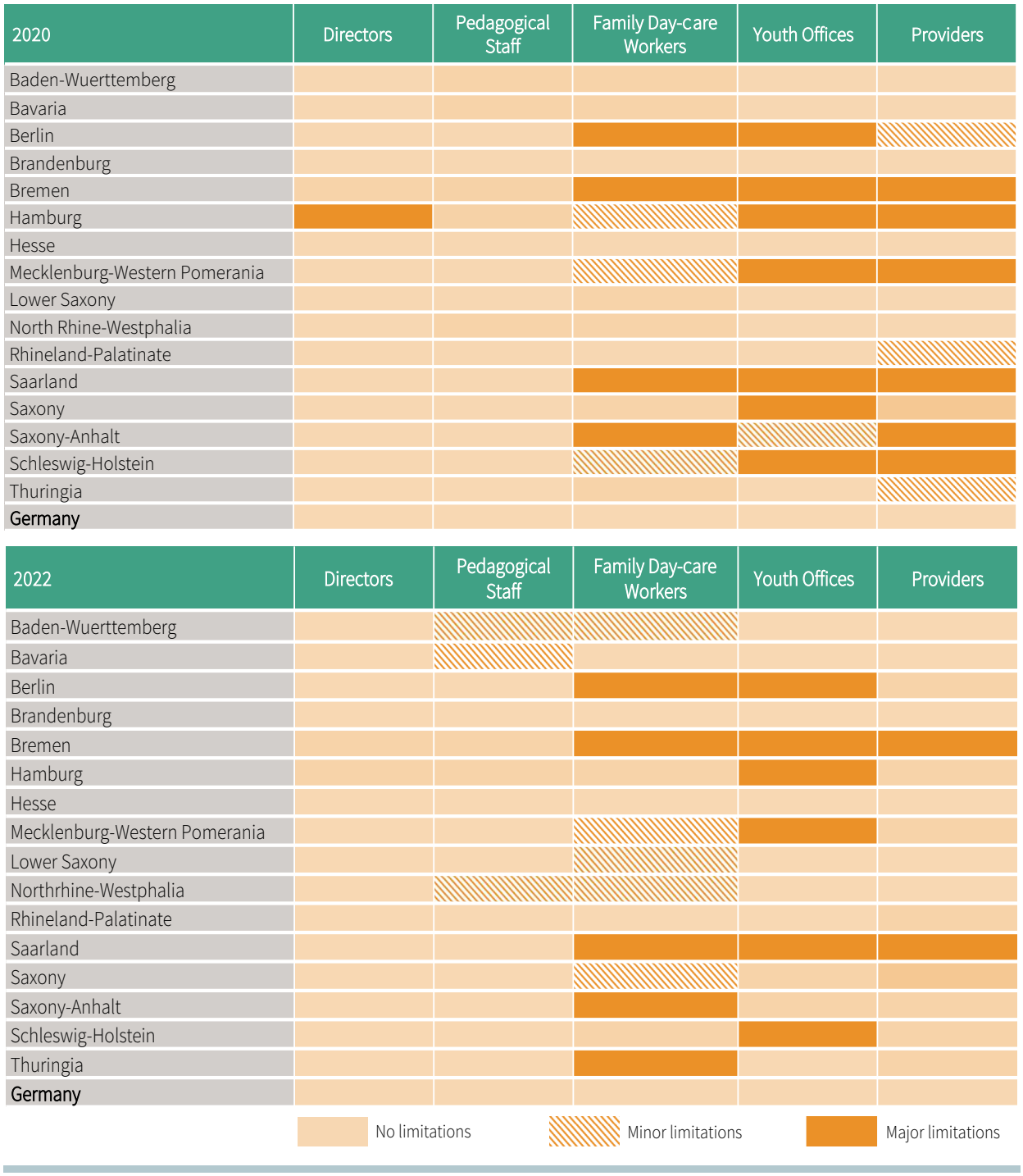
Regarding childcare providers, Bremen and Saarland experienced significant limitations solely due to low case numbers, specifically fewer than 50 childcare providers. Restrictions in local youth welfare offices were observed in Berlin, Bremen, Hamburg, Mecklenburg-Western Pomerania, Saarland, and Schleswig-Holstein, all due to insufficient case numbers. Please note that this limitation is unavoidable in many smaller states where only a few youth welfare offices exist. Consequently, the response rates of youth offices can generally be considered relatively high; nevertheless, the case numbers in some states are often too small to make valid statements about the entire population.

For the children's survey, a multi-stage procedure was applied, recruiting parents through the day-care centres and children through the parents. The day-care centres were drawn from the pool of facilities that had already participated in the ERiK directors and pedagogical staff survey 2020. Only centres where both the director and at least one pedagogical employee had participated were considered (n=2,211 facilities). The selection of the 550 centres for the pilot sample of the children and parent survey was conducted through a stratified random sampling based on regional characteristics (see Subsection 3.1 or von der Burg et al. (2021)). This approach leads to a highly selective sample in the children's survey, and the sample sizes in individual federal states are very small. Therefore, making statements based on the children's data is advisable only at the national level.

Samples of directors, pedagogical staff and childcare providers are sufficiently sized in most federal states

The current research report now implements the following regulation, which is also suggested to data users: In contrast to 2020, federal states with minor restrictions can also be evaluated and interpreted in 2022. Minor restrictions can only be found for family day-care workers and pedagogical staff. In the case of family day-care workers, the overall response rate was low; in the case of educational staff, significantly fewer educational staff could be reached than expected given the size of the target population, particularly in large federal states. In both cases, this was probably

Figure 4.4-1: **Informative Value of the ERiK Surveys at the Federal and State Levels for the years 2020 and 2022**



due to the indirect sampling and contacting of the target groups (family day-care workers were contacted via the responsible youth welfare offices and pedagogical staff via the directors of their institution, see also Section 3.3). As a result, a possible self-selection of participants in the analyses should be taken into account for the surveys and federal states that were marked with minor restrictions.

If there are severe restrictions in a federal state in one of the two survey years, comparisons between the years should not be made. The same applies if there are minor restrictions in both survey years or in 2020. Time comparisons can be made if there are either no restrictions in both survey years or only minor restrictions can be identified in 2022 (see Figure 4.4-1).

5. Summary

The primary objective of this report was to describe the methodological concept of data collection and evaluation in the ERiK Surveys 2022. Significant aspects such as the sampling strategy, response rates, weighting methods and selectivity were carefully analysed. Additional methodological details on the ERiK Surveys 2020 are available in the ERiK Methodological Reports I and II (Schacht et al. 2021, 2022), and for the ERiK Surveys 2022, in the ERiK Methodological Report III (Schacht et al. 2023).

Based on the available data, general statements about the framework conditions in child day-care are possible from various perspectives. Readers of the ERiK Research Report IV and users of the data of the ERiK Surveys 2022 should consider the following points:

- In the ERiK Research Report IV (Fackler et al. 2024), generalised statements about the target groups are made with the help of the ERiK Surveys 2022. This is possible by weighting the data and taking into account the complex sample design. It is recommended to evaluate the data by taking absolute case numbers and the dispersion of the data into account for all analyses.
- Due to changes in the target populations of pedagogical staff and directors in 2022 compared to 2020, retrospective adjustments to the target populations and weights for 2020 were necessary. This may lead to different results compared to previous research reports or analyses.
- The reliability of the ERiK Surveys 2022 is limited with regard to estimates on a point and population basis in certain federal states and for specific surveys. Despite the comparatively large sample sizes of the ERiK Surveys 2022, the samples are small in some cases, especially at the federal state level. There is also the possibility of a correlation between the people who did not take part in the survey and the KiQuTG parameters under investigation. This is due to the lack of complete sampling frames in the field of ECEC and the fact that official statistics are not available for all populations to a sufficient extent.
- The ERiK survey data for 2020 is available from the DJI Research Data Center (FDZ-DJI, www.surveys.dji.de). The data for the survey year 2022 and an update to the 2020 data are expected to be made available in the course of 2024.

Overall, the ERiK Surveys provide high-quality and unique data, offering a comprehensive foundation for a wide range of in-depth and multi-perspective analyses in the field of early childhood education and care. This rich dataset enables researchers to explore various aspects of ECEC from different angles, contributing to a deeper understanding of the sector and informing evidence-based decision-making.

Finally, the following study synopsis summarises some basic information for the ERiK-Surveys 2022 (Table 5.0-1).

Table 5.0-1: **Study Synopsis ERiK-Surveys 2022**

Synopsis 2022	
Target Population	Youth welfare offices (YWO), family day-care workers (FDW), childcare providers (PRO) of and directors (DIR), pedagogical staff (PST) & children (CHI) in child day-care centres (CEN)
Field times main survey	› FDW, YWO, PRO: January - March 2022 › CEN/DIR, PST: February - April 2022 › CHI: May - September 2022
Target/interviewees	› CEN/DIR: persons with the highest share of managerial tasks in the centre › PST/FDW: selected persons › YWO/PRO: heads of department, deputy heads, one or more staff members › CHI: children aged 4-7 in a day-care centre, not attending school yet
Survey institutes	› CEN/DIR, PST, CHI: infas Institute for Applied Social Sciences › FDW, YWO, PRO: SOKO Institute for Social Research and Communication
Survey instruments	› Paper postal self-completion and access to the online version of the questionnaire (P&O) › Access to the online version of the questionnaire (O) › In-person interview in the centre (only children)
Contact and contact possibility for target persons	› Target population-specific cover letter with information on study and data protection, postal reminder and telephone reminder (incl. nonresponse survey (for PRO, DIR)) › Information telephone number and study-specific e-mail address of the survey institutes
Gross sample (GS), Achieved sample size/net cases (NC) and AAPOR response rate 2 (RR)	› L: GS: 18,000; NC: 4,832, of which 4,674 complete (RR: 27 %) › P: GS: 40,800; NC: 7,116, of which 7,019 complete (RR: 18 %) › K: GS: 43,800 (complete population survey); NC: 3,927, of which 3,854 complete (RR: 9 %) › J: GS: 569/575 (complete population survey); NC: 366, of which 341 complete (RR: 64 %) › T: GS: 21,600 (complete population survey); NC: 5,166, of which 4,710 complete (RR: 24 %) › C: GS: 712; NC: 490, of which 479 complete (RR: 69 %)
Weighting	› Design weighting for CEN/DIR, PST, FDW, CHI › Nonresponse weighting using random forest models for DIR, PST, YWO, PRO, CHI › Calibration/adjustment weighting for CEN, FDW, YWO, PRO, CHI
Citation of data	› Total dataset: Gedon, Benjamin/Schacht, Diana D./Gilg, Jakob J./Classe, Franz L./Herrmann, Sonja/Brusis, Martin/Buchmann, Janette/Drexl, Doris/Guck, Christian/Kuger, Susanne/Müller, Michael/Preuß, Melina/Romefort, Johanna/Ulrich, Lisa/Wenger, Felix (2023): ERiK-Surveys 2022. Deutsches Jugendinstitut (DJI). München. Datensatz Version 2.0. https://doi.org/10.17621/erik2022_v02 › Survey specific datasets (see Bibliography): **DIR**: Gedon et al. (2023b); **PST**: Gedon et al. (2023a); **FDW**: Gedon et al. (2023c); **YWO**: Gedon et al. (2023d); **PRO**: Gedon et al. (2023e); **CHI**: Maron et al. (2023).

Note: Abbreviations: CEN = Centres, DIR = Directors, PST = Pedagogical Staff, FDW = Family Day-Care Workers, YWO = Youth Welfare Offices, PRO = Providers, CHI = Children.

Bibliography

Bertelsmann Stiftung (2024): Wegweiser Kommune 2022. Kommunale Daten für eine innovative Zukunft. URL: https://www.wegweiser-kommune.de/ (11/06/2024)

Bethlehem, Jelke (2009): Applied survey methods: a statistical perspective. Hoboken, NJ

Breiman, Leo (2001): Random forests. In: Machine learning, 45. Vol. Pp. 5–32

Buchmann, Janette/Ziesmann, Tim/Drexl, Doris (2022): HF-04 Stärkung der Leitung. In: Klinkhammer, Nicole/Schacht, Diana D./Meiner-Teubner, Christiane/Kuger, Susanne/Kalicki, Bernhard/Riedel, Birgit (Eds.): ERiK-Forschungsbericht II. Befunde des indikatorengestützten Monitorings zum KiQuTG. Bielefeld, pp. 109–123

Bundesinstitut für Bau-, Stadt- und Raumforschung (2022): INKAR – Indikatoren und Karten zur Raum- und Stadtentwicklung. Indikatorenübersicht. URL: https://www.inkar.de/documents/Uebersicht%20der%20Indikatoren.xlsx (11/06/2024)

Buskirk, Trent D./Kolenikov, Stanislav (2015): Finding respondents in the forest: a comparison of logistic regression and random forest models for response propensity weighting and stratification. In: Survey Insights: Methods from the Field, Weighting: Practical Issues and 'How to' Approach, pp. 1–17

Deming, W. Edwards/Stephan, Frederick F. (1940): On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. In: The Annals of Mathematical Statistics, 11. Vol., Issue 4, pp. 427–444

European Commission (2021): Country compendium. A companion to the english style guide

Fackler, Sina/Herrmann, Sonja/Meiner-Teubner, Christiane/Bopp, Christine/Kuger, Susanne/Kalicki, Bernhard (Eds.) (2024): ERiK-Forschungsbericht IV. Befunde des indikatorengestützten Monitorings zum KiQuTG. Bielefeld

Federal Statistical Office (2019): Statistisches Jahrbuch. Deutschland und Internationales

Federal Statistical Office (2022): Statistiken der Kinder- und Jugendhilfe. Kinder und tätige Personen in Tageseinrichtungen und in öffentlich geförderter Kindertagespflege am 01.03.2022. Wiesbaden

Ferri-García, Ramón/Beaumont, Jean-François/Bosa, Keven/Charlebois, Joanne/Chu, Kenneth (2022): Weight smoothing for nonprobability surveys. In: TEST, 31. Vol., Issue 3, pp. 619–643

Ferri-García del Mar (2020): Propensity score adjustment using machine learning classification algorithms to control selection bias in online surveys. In: PloS one, 15. Vol., Issue 4, e0231500

Gedon, Benjamin/Schacht, Diana D./Gilg, Jakob J./Classe, Franz L./Herrmann, Sonja/Brusis, Martin/Buchmann, Janette/Guck, Christian/Kuger, Susanne/Pachner, Theresia/Romefort, Johanna/Wenger, Felix (2023a): ERiK-Surveys 2022: Befragung pädagogisches Personal. Deutsches Jugendinstitut (DJI). Datensatz Version 2.0. https://doi.org/10.17621/erik2022_p_v02

Gedon, Benjamin/Schacht, Diana D./Gilg, Jakob J./Classe, Franz L./Herrmann, Sonja/Buchmann, Janette/Drexl, Doris/Kuger, Susanne/Pachner, Theresia (2023b): ERiK-Surveys 2022: Leitungsbefragung. Deutsches Jugendinstitut (DJI). Datensatz Version 2.0. https://doi.org/10.17621/erik2022_l_v02

Gedon, Benjamin/Schacht, Diana D./Gilg, Jakob J./Classe, Franz L./Herrmann, Sonja/Kuger, Susanne/Müller, Michael/Tursun, Nadira (2023c): ERiK-Surveys 2022: Befragung Kindertagespflegepersonen. Deutsches Jugendinstitut (DJI). Datensatz Version 2.0. https://doi.org/10.17621/erik2022_k_v02

Gedon, Benjamin/Schacht, Diana D./Gilg, Jakob J./Classe, Franz L./Herrmann, Sonja/Kuger, Susanne/Preuß, Melina (2023d): ERiK-Surveys 2022: Jugendamtsbefragung. Deutsches Jugendinstitut (DJI). Datensatz Version 2.0. https://doi.org/10.17621/erik2022_j_v02

Gedon, Benjamin/Schacht, Diana D./Gilg, Jakob J./Classe, Franz L./Herrmann, Sonja/Kuger, Susanne/Ulrich, Lisa (2023e): ERiK-Surveys 2022: Trägerbefragung. Deutsches Jugendinstitut (DJI). Datensatz Version 2.0. https://doi.org/10.17621/erik2022_t_v02

Gelein, Brigitte/Haziza, David/Causeur, David (2018): Propensity weighting for survey nonresponse through machine learning. In: 13es Journées de méthodologie statistique de l'Insee (JMS)

Groves, Robert M. (2004): Survey errors and survey costs. Wiley series in survey methodology. Vol. 581. New York

Groves, Robert M./Lyberg, Lars (2010): Total survey error: Past, present and future. In: Public Opinion Quarterly, 74. Vol., Issue 5, pp. 849–879

Hastie, Trevor/Tibshirani, Robert/Friedman, Jerome H. (2009): The elements of statistical learning: data mining, inference, and prediction. Vol. 2

Kauermann, Göran/Küchenhoff, Helmut (2010): Stichproben. Methoden und praktische Umsetzung mit R. Springer-Lehrbuch. Berlin/ Heidelberg

Kern, Christoph/Klausch, Thomas/Kreuter, Frauke (2019): Tree-based machine learning methods for survey research. In: Survey research methods. Vol. 13. NIH Public Access, p. 73

Klinkhammer, Nicole/Schacht, Diana D./Meiner-Teubner, Christiane/Kuger, Susanne/Kalicki, Bernhard/Riedel, Birgit (Eds.) (2022): ERiK-Forschungsbericht II. Befunde des indikatorengestützten Monitorings zum KiQuTG. Bielefeld

Kruskal, William/Mosteller, Frederick (1980): Representative sampling IV. In: International Statistical Review, 48. Vol., Issue 2, pp. 169–195

Lee, Brian K./Lessler, Justin/Stuart, Elizabeth A. (2010): Improving propensity score weighting using machine learning. In: Statistics in Medicine, 29. Vol., Issue 3, pp. 337–346

Lumley, Thomas (2023): Analysis of complex survey samples. Version R package version 4.2-1. URL: http://r-survey.r-forge.r-project.org/survey/

Maron, Julian/Klug, Christina/Leßner, Lisa/Classe, Franz/Gilg, Jakob J./Pachner, Theresia/Rahmann, Susanne/Schacht, Diana D./Molina Ramirez, Magdalena/Fichter, Magdalena/Kuger, Susanne (2023): ERiK-Surveys 2022: Kinderbefragung. Deutsches Jugendinstitut (DJI). Datensatz Version 1.0. https://doi.org/10.17621/erik2022_c_v01

Probst, Philipp/Wright, Marvin N./Boulesteix, Anne-Laure (2019): Hyperparameters and tuning strategies for random forest. In: Wiley Interdisciplinary Reviews: data mining and knowledge discovery, 9. Vol., Issue 3, e1301

Research Data Centre of the Statistical Offices of the Federal States (2022): Statistiken der Kinder- und Jugendhilfe, On-Site, Version 0. Research Data Centre of the Statistical Offices of the Federal States. URL: https://www.forschungsdatenzentrum.de/de/kinder-und-jugendhilfe/kjh (05/08/2022)

Research Data Centre of the Statistical Offices of the Federal States (2024): Statistiken der Kinder- und Jugendhilfe, On-Site. Research Data Centre of the Statistical Offices of the Federal States. URL: https://www.forschungsdatenzentrum.de/de/kinder-und-jugendhilfe/kjh (23/01/2024)

Sand, Matthias/Kunz, Tanja (2020): Gewichtung in der Praxis. Mannheim

Schacht, Diana D./Gedon, Benjamin/Gilg, Jakob J./Klug, Christina/Kuger, Susanne (2021): ERiK-Methodological Report I. Target populations, sampling frame and sampling design of the ERiK-Surveys 2020. Bielefeld

Schacht, Diana D./Gedon, Benjamin/Gilg, Jakob J./Kuger, Susanne (2022): ERiK-Methodological Report II. Implementation, data quality and data structure of the ERiK-Surveys 2020. Bielefeld

Schacht, Diana D./Gilg, Jakob J./Gedon, Benjamin/Brusis, Martin/Buchmann, Janette/Drexl, Doris/Hoang, Tony/Jähnert, Alexandra/Kuger, Susanne/Molina Ramirez, Magdalena/Müller, Michael/Preuß, Melina/Rahmann, Susanne/Ulrich, Lisa/Wenger, Felix (2023): ERiK-Methodological Report III. Conceptualisation and design of the ERiK-Surveys 2022. Bielefeld

Schacht, Diana D./Ulrich, Lisa (2024): Befragungen von Trägern von Kindertageseinrichtungen: Eine Methodenstudie zu Verzerrungen. In: Zeitschrift für Erziehungswissenschaft, 27. Vol. Pp. 497–516

Stadtmüller, Sven/Silber, Henning/Daikeler, Jessica/Martin, Silke/Sand, Matthias/Schmich, Patrick/Schröder, Jette/Struminskaya, Bella/Weyandt, Kai W./Zabal, Anouk (2019): Adaptation of the AAPOR final disposition codes for the german survey context. Mannheim

Valliant, Richard/Dever, Jill A./Kreuter, Frauke (Eds.) (2013): Practical tools for designing and weighting survey samples. Chapter 15: Variance Estimation. 2nd ed. New York

von der Burg, Julian/Eibelshäuser, Svenja/Kleudgen, Martin/Torregroza, Sabrina (2021): Methodenbericht: Entwicklung von Rahmenbedingungen der Kindertagesbetreuung (ERiK) - Kinder- und Elternbefragung. Bonn

Watkins, Stephanie/Jonsson-Funk, Michele/Brookhart, M. Alan/Rosenberg, Steven A./O'Shea, T. Michael/Daniels, Julie (2013): An empirical comparison of tree-based methods for propensity score estimation. In: Health services research, 48. Vol., Issue 5, pp. 1798–1817

Westreich, Daniel/Lessler, Justin/Funk, Michele Jonsson (2010): Propensity score estimation: neural networks, support vector machines, decision trees (CART), and meta-classifiers as alternatives to logistic regression. In: Journal of clinical epidemiology, 63. Vol., Issue 8, pp. 826–833

Wieschke, Gustav Johannes/Lippert, Kerstin/Kuger, Susanne (2023): Methodischer Hintergrund der KiBS-Erhebung 2022. DJI-Kinderbetreuungsreport 2023, Studie 7 von 7. München

Zaloznik, Maja (2011): Iterative proportional fitting - theoretical synthesis and practical limitations. Liverpool

A. Appendix

Figure A.0-1: **Question on Gender in the ERiK-Survey 2022 of Pedagogical Staff**

42 Welches Geschlecht haben Sie?

☞ *Bitte machen Sie nur **eine** Angabe.*

Männlich 1 ☐ Weiblich 2 ☐ Divers 3 ☐

Figure A.0-2: **Extract from the letter to the directors about which employees should take part in the ERiK surveys**

Was hat es mit den beiliegenden Umschlägen für „pädagogisches Personal" auf sich?

Die ERiK-Studie sammelt Daten und Einschätzungen aus ganz unterschiedlichen Perspektiven. Für uns sind daher auch die Angaben des pädagogischen Personals wichtig. Bitte übergeben Sie daher die beiliegenden **Umschläge für das pädagogische Personal** an [bdx] Ihrer pädagogisch tätigen MitarbeiterInnen. Für die Aussagekraft der Studie ist es wichtig, dass die Auswahl des pädagogischen Personals möglichst zufällig erfolgt. Um die TeilnehmerInnen für die Befragung möglichst zufällig auszuwählen, geben Sie bitte die Umschläge möglichst an die [bdx] KollegInnen, die zuletzt Geburtstag hatten.

Diese KollegInnen

- **müssen** pädagogisch tätig sein, aber
- dürfen **nicht** als Einrichtungsleitung beschäftigt sein und
- dürfen **nicht** ehrenamtlich in Ihrer Einrichtung tätig sein.

Aus datenschutzrechtlichen Gründen nehmen Sie bitte keine Einsicht in ausgefüllte Fragebögen des pädagogischen Personals.

Figure A.0-3: **Extract from the questionnaire for pedagogical staff**

1 Sind Sie in Ihrer Einrichtung als Einrichtungsleitung tätig?

☞ *„Einrichtungsleitung" umschließt auch die Positionen „stellvertretende Leitung", „Teil eines Leitungsteams" oder „kommissarische Leitung".*

Ja 1 ☐ → ***Leider gehören Sie nicht zur Zielgruppe unserer Befragung. Bitte schicken Sie den Fragebogen dennoch an uns zurück.***

Nein 2 ☐ → ***Bitte weiter mit Frage 2***

Table A.0-1: **Unweighted and weighted data for pedagogical staff for the survey year 2022**

Mean of contractually agreed weekly working hours per birth cohort				
	Unweighted		Weighted	
	Estimate	S.E.	Estimate	S.E.
1945-1960	32.252	0.503	32.323	0.493
1961-1970	32.764	0.180	33.226	0.178
1971-1980	32.449	0.172	32.527	0.171
1981-1990	33.256	0.138	33.745	0.146
1991-2003	35.915	0.124	36.946	0.108
N	6,911		6,839	

Table A.0-2: **Unweighted and weighted data for pedagogical staff for the survey year 2020**

Mean of contractually agreed weekly working hours per birth cohort						
	Unweighted		Weighted without Calibration		Weighted with Calibration	
	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.
1945-1960	31.900	0.298	31.508	0.337	31.296	0.299
1961-1970	32.457	0.159	31.937	0.183	31.649	0.168
1971-1980	31.061	0.171	30.435	0.191	30.137	0.177
1981-1990	33.094	0.156	32.823	0.179	32.513	0.164
1991-2003	36.445	0.120	36.621	0.129	35.832	0.146
N	8,289		6,930		6,930	

Table A.0-3: **Unweighted and weighted data for directors for the survey year 2022**

Mean of contractually agreed weekly working hours per birth cohort				
	Unweighted		Weighted	
	Estimate	S.E.	Estimate	S.E.
1945-1960	36.537	0.150	37.679	0.129
1961-1970	36.761	0.060	37.209	0.068
1971-1980	36.135	0.075	36.600	0.081
1981-1990	36.226	0.077	36.842	0.092
1991-2003	37.355	0.113	37.910	0.115
N	15,140		9,688	

Table A.0-4: **Unweighted and weighted data for directors for the survey year 2020**

Mean of contractually agreed weekly working hours per birth cohort						
	Unweighted		Weighted without Calibration		Weighted with Calibration	
	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.
1945-1960	36.791	0.195	36.699	0.200	36.221	0.228
1961-1970	36.663	0.120	36.601	0.121	36.106	0.133
1971-1980	35.878	0.158	35.571	0.169	35.016	0.177
1981-1990	36.356	0.210	36.471	0.203	35.842	0.228
1991-2003	38.098	0.204	38.361	0.182	38.149	0.207
N	3,707		3,707		3,707	

Full List of Regional Variables used in the 2022 Random Forest Models:

District Level Data

- 2020 Beschäftigungsanteil im 2. Sektor (%)
- 2020 Beschäftigungsanteil im 3. Sektor (%)
- 2020 Beschäftigungsquote (%)
- 2020 Geringfügig Beschäftigte (Wohnort) (je 1.000 Einwohner:innen)
- 2020 Bevölkerung (Anzahl)
- 2020 Bevölkerungsentwicklung über die letzten 5 Jahre (%)
- 2020 Geburten (je 1.000 Einwohner:innen)
- 2020 Anteil Elternjahrgänge (%)
- 2020 Anteil 65- bis 79-Jährige (%)
- 2020 Anteil ab 80-Jährige (%)
- 2020 Einwohner:innendichte (Einwohner:innen je Hektar)
- 2020 Steuereinnahmen pro Einwohner:in (Euro je Einwohner:in)
- 2020 Soziale Leistungen (Euro je Einwohner:in)
- 2020 Jugendhilfe (Euro je Einwohner:in)
- 2020 SGB II-/SGB XII-Quote (%)
- 2020 Kinderarmut (%)
- 2020 Jugendarmut (%)
- 2020 Altersarmut (%)
- 2020 Unter 3-Jährige in Tageseinrichtungen (%)
- 2020 Langzeitarbeitslosenquote (%)
- 2020 Unter 3-Jährige in Tageseinrichtungen – 25 bis 35 h Betreuung (%)
- 2020 Unter 3-Jährige in Tageseinrichtungen – mehr als 35 h Betreuung (%)
- 2020 3- bis 5-Jährige in Tageseinrichtungen (%)
- 2020 3- bis 5-Jährige in Tageseinrichtungen – 25 bis 35 h Betreuung (%)
- 2020 3- bis 5-Jährige in Tageseinrichtungen – mehr als 35 h Betreuung (%)
- 2020 Personal mit Hochschulabschluss in Tageseinrichtungen (%)
- 2020 Unter 3-Jährige in Tagespflege – bis 25 h Betreuung (%)
- 2020 Unter 3-Jährige in Tagespflege – mehr als 35 h Betreuung (%)
- 2020 3- bis 5-Jährige in Tagespflege (%)
- 2020 3- bis 5-Jährige in Tagespflege – bis 25 h Betreuung (%)
- 2020 3- bis 5-Jährige in Tagespflege – 25 bis 35 h Betreuung (%)
- 2020 3- bis 5-Jährige in Tagespflege – mehr als 35 h Betreuung (%)
- 2020 6- bis 10-Jährige in Tagespflege (%)
- 2020 Ausländer:innen (Einwohner:innen)
- 2020 Anteil Ausländer:innen (%)
- 2020 3-Jährige mit Migrationshintergrund in Tageseinrichtungen (%)
- 2020 Kinder mit Migrationshintergrund in Tageseinrichtungen (%)
- 2020 Arbeitslose an der Gesamtbevölkerung (%)
- 2020 SGB II-Quote (%)
- 2020 Schulabgänger:innen mit Förderschulabschluss – Gesamt (%)
- 2020 Schulabgänger:innen mit Realschulabschluss – Gesamt (%)
- 2020 Schulabgänger:innen allgmb./berufsb. Schulen mit Fachhoch-/Hochschulreife – Gesamt (%)
- 2020 ALG II-Quote (%)
- 2020 Bevölkerung 0- bis 2-Jährige (Einwohner:innen)
- 2020 Bevölkerung 3- bis 5-Jährige (Einwohner:innen)
- 2020 Bevölkerung 6- bis 9-Jährige (Einwohner:innen)
- 2020 Entwicklung 3- bis 5-Jährige seit 2011 (%)

Municipality Level Data

- Bevölkerung
- Bodenfläche gesamt qkm
- Einwohnerdichte
- Bevölkerungsentwicklung
- Siedlungs- und Verkehrsfläche
- Siedlungsdichte in km²
- Gerichte
- Amtsgerichte
- Nebenstelle eines Amtsgerichts
- Landgerichte
- Nebenstelle eines Landgerichts
- Oberlandesgerichte
- Nebenstelle eines Oberlandesgerichts
- sozialversicherungspflichtig Beschäftigte am Arbeitsort
- sozialversicherungspflichtig Beschäftigte am Wohnort
- Beschäftigtendichte (AO)
- Beschäftigtendichte (WO)
- sozialversicherungspflichtig Beschäftigte Einpendler
- sozialversicherungspflichtig Beschäftigte Auspendler

- sozialversicherungspflichtig Beschäftigte Binnenpendler
- Pendlersaldo
- Arbeitsplatzzentralität
- Arbeitslosigkeit
- ÖV-Haltestellen
- Bahnhaltestellen
- Bushaltestellen
- U-/Strassenbahnhaltestellen
- hochfrequentierte ÖV-Haltestellen
- hochfrequentierte Bahnhaltestellen
- hochfrequentierte Bushaltestellen
- hochfrequentierte U-/Strassenbahnhaltestellen
- ÖV-Abfahrten
- Bahn-Abfahrten
- Bus-Abfahrten
- U-/Straßenbahn-Abfahrten
- Erreichbarkeit von IC-/EC-/ICE-Bahnhöfen
- Erreichbarkeit Autobahnen
- Erreichbarkeit Flughäfen
- Erreichbarkeit KV-Terminal
- Erreichbarkeit von Mittelzentren
- Erreichbarkeit von Oberzentren
- Breitbandversorgung mit 1000 Mbit/s in
- Ärzte
- Einwohner je Arzt
- Hausärzte
- Allgemeinärzte
- Internisten
- Kinderärzte
- Hochschulen gesamt
- Studierende an Hochschulen
- Studierende an Hochschulen je 1.000 Einwohner
- Studierende an Fachhochschulen je 1.000 Einwohner
- Zentralörtlicher Status (kategorial)
- Zentralörtlicher Status (zusammengefasst) (kategorial)
- Stadt-/Gemeindetyp (kategorial)
- Raumtyp nach Lage (kategorial)
- RegioStaR 17 – Regionalstatistischer Raumtyp (kategorial)

Overview of ERiK Weights 2020:

Directors:

- `nwd`: Design weight
- `nwn`: Non-response weight (Logit)
- `nww`: Uncalibrated total weight at the respondent level
- `nww_kal`: Calibrated total weight at the respondent level (calibrated to KJH Directors)
- `nwe`: Calibrated total weight at the centre level, trimmed

Pedagogical staff:

- `nwd`: Design weight
- `nwd_22`: Design weight
- `nwn`: Non-response weight (Logit)
- `nwn_22`: Non-response weight (Logit)
- `nww`: Uncalibrated total weight for 2020 Def.
- `nww_22`: Uncalibrated total weight for 2022 Def.
- `nww_kal`: Calibrated total weight for 2020 Def.

Providers:

- `nww`: Trimmed total weight at the provider level
- `nwe`: Trimmed total weight at the institutional level
- `nww_01`: Untrimmed total weight at the provider level
- `nwe_01`: Untrimmed total weight at the institutional level

Youth Welfare Offices:

- `nwn`: Nonresponse weight
- `nww`: Trimmed total weight at the respondent level

Family Day-Care Workers:

- `nwd`: Design Weight
- `nww`: Trimmed total weight at the respondent level

Overview of ERiK Weights 2022:

Directors:

- `nwd`: Design weight
- `nwn`: Nonresponse weight (Random Forest)
- `nww`: Uncalibrated total weight at the respondent level, trimmed

- `nww_01:` Uncalibrated total weight at the respondent level, untrimmed
- `nwe:` Calibrated total weight at the centre level, trimmed
- `nwe_01:` Calibrated total weight at the centre level, untrimmed

Pedagogical staff:

- `nwd:` Design weight
- `nwn:` Nonresponse weight (Random Forest)
- `nww:` Uncalibrated total weight, trimmed
- `nww_01:` Uncalibrated total weight, untrimmed

Providers:

- `nww:` Trimmed total weight at the provider level
- `nwe:` Trimmed total weight at the insitutional level
- `nww_01:` Untrimmed total weight at the provider level
- `nwe_01:` Untrimmed total weight at the insitutional level

Youth Welfare Offices:

- `nwn:` Nonresponse weight
- `nww:` Trimmed total weight at the respondent level
- `nww_01:` Untrimmed total weight at the respondent level

Family Day-Care Workers:

- `nwd:` Nonresponse weight
- `nww:` Trimmed total weight at the respondent level
- `nww_01:` Untrimmed total weight at the respondent level

Children/Parents:

- `nww:` Trimmed total weight at the respondent level
- `nww_01:` Untrimmed total weight at the respondent level

List of Figures

List of Tables

ERiK-Methodological Report IV

Implementation and Data Quality of the ERiK-Surveys 2022

The ERiK-Methodological Report IV is the fourth in a series of methodological reports related to the 'Entwicklung von Rahmenbedingungen in der Kindertagesbetreuung – indikatorengestützte Qualitätsbeobachtung (ERiK)' study. The report introduces the implementation of the sample and survey designs and evaluates the data quality of the ERiK-Surveys 2022. Together with the ERiK-Methodological Report III, it contains all background information on the ERiK-Surveys 2022.

Researching children, youth and families at the intersection of science, policy, and professional practice

The German Youth Institute (DJI) is one of the largest social science research institutes in Europe with an experience of over 60 years. The DJI conducts empirical studies into the life situations of children, young people and families, and provides policy advice to the German national government, the German federal states and local authorities as well as key impulses for professional practice.

Authors:
Sonja Herrmann, Jakob J. Gilg, Franz L. Classe, Diana D. Schacht, Benjamin Gedon, Susanne Kuger

wbv

ISBN: 978-3-7639-7784-0