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

# FDZ-METHODENREPORT

Methodological aspects of labour market data

## 07|2025 EN Linking AID:A survey data and administrative labor market biography data from the Institute for Employment Research (AID:A-ADIAB 7523)

Manfred Antoni, Christina Boll, Sandra Dummert, Simone Schüller

Bundesagentur für Arbeit

# Linking AID:A survey data and administrative labor market biography data from the Institute for Employment Research (AID:A-ADIAB 7523)

Manfred Antoni (Institute for Employment Research, IAB),
Christina Boll (German Youth Institute, DJI),
Sandra Dummert (IAB),
Simone Schüller (DJI)

Die FDZ-Methodenreporte befassen sich mit den methodischen Aspekten der Daten des FDZ und helfen somit Nutzerinnen und Nutzern bei der Analyse der Daten. Nutzerinnen und Nutzer können hierzu in dieser Reihe zitationsfähig publizieren und stellen sich der öffentlichen Diskussion.

FDZ-Methodenreporte (FDZ method reports) deal with methodical aspects of FDZ data and help users in the analysis of these data. In addition, users can publish their results in a citable manner and present them for public discussion.

# Contents

# Abstract

This paper documents the process of linking the DJI survey Growing up in Germany (AID:A) with administrative labor market biography data from the Institute for Employment Research (IAB). Among AID:A-2023 respondents aged 16 and above, the linkage consent rate was 82.9%, which exceeded *a priori* expectations. The success rate of the linkage was 95.7%, which is above the average of previous linkage projects with IAB data. We explore linkage consent bias as well as linkage success bias conditional on consent.

# Zusammenfassung

Dieses Papier dokumentiert den Prozess der Verknüpfung der DJI-Studie “Aufwachsen in Deutschland: Alltagswelten” (AID:A) mit administrativen Erwerbsverlaufsdaten des Instituts für Arbeitsmarkt- und Berufsforschung (IAB). Unter den AID:A-2023-Befragten im Alter von 16 Jahren und älter lag die Zustimmungsrate zur Datenverknüpfung bei 82,9% und übertraf damit die initialen Erwartungen. Die Erfolgsrate der Verknüpfung betrug 95,7% und liegt damit über dem Durchschnitt früherer Verknüpfungsprojekte mit IAB-Daten. Es werden zudem mögliche Verzerrungen in der Verknüpfungseinwilligung sowie im Verknüpfungserfolg (unter den Einwilligenden) untersucht.



# Keywords

administrative data, AID:A, employment data, Germany, record linkage, survey data

# Acknowledgements

This record linkage project was funded by the German Federal Ministry for Education, Family Affairs, Senior Citizens, Women and Youth (*Bundesministerium für Bildung, Familie, Senioren, Frauen und Jugend*, BMBFSFJ) under grant number 3023DJIRLI. We thank Lisa Schmidtlein and Steffen Kaimer from IAB’s department Data and IT-Management for providing the linkage identifiers from the data warehouse of the Federal Employment Agency, and Dominik Braun and Katrina Blindow (both DJI) for helpful suggestions.

# 1 Introduction

This working paper documents the record linkage of the DJI survey Growing up in Germany (*Aufwachsen in Deutschland: Alltagswelten*, AID:A) with administrative labor market biography data from the Institute for Employment Research (*Institut für Arbeitsmarkt- und Berufsforschung*, IAB) of the German Federal Employment Agency (*Bundesagentur für Arbeit*, BA). The IAB did perform the record linkage on behalf of the German Youth Institute (*Deutsches Jugendinstitut*, DJI). The linkage project was funded by the German Federal Ministry for Education, Family Affairs, Senior Citizens, Women and Youth (*Bundesministerium für Bildung, Familie, Senioren, Frauen und Jugend*, BMBFSFJ).

During the 2023 wave of AID:A, respondents aged 16 and above were asked to provide consent for data linkage, resulting in an overall consent rate of 82.9%, which surpassed initial expectations. Our record linkage achieved a remarkable success rate of 95.7%, exceeding the average of previous linkage projects using IAB data. We further investigate potential biases (consent bias and success bias) and identify respondent age as a relevant factor to be taken into account in subsequent analyses.

As this paper focuses on the specific challenges and decisions in the context of the current linkage project, we do not provide a comprehensive overview of linkage techniques. For an in-depth coverage of the methods used here, see, for instance, Christen (2012). We use a set of programs developed by the German Record Linkage Center (GRLC, see https://record-linkage.de). Antoni/Schnell (2019) provide a general overview of the GRLC's activities.

Contractual restrictions and capacity constraints at the Research Data Centre (FDZ) of the German Federal Employment Agency (BA) at the IAB prevent the direct release of the resulting data to the general scientific community as an FDZ data product. Researchers interested in using the linked data are invited to contact the authors affiliated with the DJI to explore a potential collaboration. Even in the case of a collaboration with the DJI, access to the linked data would be fee-based.

The remainder of this paper is structured as follows: Section 2 outlines the procedure and results of obtaining informed consent for data linkage among AID:A respondents and compares the characteristics of those who provided consent with those who did not. Section 3 describes the two data sources and the non-unique linkage identifiers they contained. Section 4 provides information on the cleaning and standardizing of these identifiers. Section 5 describes the comparison and classification steps of the linkage process. Section 6 discusses the linkage results, and compares characteristics of matched and unmatched survey respondents. Section 7 concludes with a short summary and some final remarks.

# 2 Linkage consent

## 2.1 Respondents' informed consent to data linkage

The linkage between the survey and administrative data is based on the AID:A respondents. Beginning at the age of 16 at the time of the interview, participants in the AID:A 2023 survey wave were asked to provide their consent for linking their survey data with the administrative data concerning them that is available at the IAB for research purposes. Respondents were informed about the linkage procedure and asked for their consent as follows:

> To keep the interview as brief as possible in the future, we would like to include extracts from data held about you at the Institute for Employment Research (IAB) in Nuremberg in the analysis of this survey. These data include, for example, additional information on periods of employment, unemployment, and participation in programs during unemployment. For the purpose of linking the data, we will transmit your name, address, date of birth, and gender to the IAB. After the linkage has been completed, these identifiers will be deleted immediately at the IAB, and only then will your survey data be transferred to the IAB. We kindly ask for your consent. When these data are analyzed, strict compliance with all data protection regulations is absolutely guaranteed. Providing your consent is, of course, voluntary. You will not experience any disadvantages if you choose not to consent, and you may withdraw your consent at any time. Do you agree to the transmission and linkage of your data?

Respondents aged 16-17 could additionally choose whether they preferred to be addressed formally *("Sie")* or informally *("du")*. In addition, the sentence *"This includes, for example, additional information on periods of employment, unemployment, and participation in programs during unemployment"* was slightly adapted to read *"This includes, for example, additional information on career counseling or employment."* In all survey modes, two information buttons providing additional details were available:

> **How can the data linkage be anonymous?** The linkage is carried out automatically using an identification code. After the linkage has been completed, this code is deleted. We guarantee that all information will be treated confidentially and analyzed without names or addresses (i.e., anonymously). All results are evaluated only at the group level.

> **Where does this information come from?** The information is provided by the social insurance institutions and may be used for analysis purposes only if you give your consent.

Already in the postal cover letter used to contact the AID:A respondents, reference was made to the upcoming consent request for data linkage:

> In the interview, respondents aged 16 and older are asked for permission to link their responses with research data from the Institute for Employment Research (IAB) of the Federal Employment Agency. Providing consent is voluntary and can be withdrawn at any time. Participation in the survey is independent of this consent, and refusal will not result in any disadvantages. The IAB research data include information on your employment history, job search, work relationships, establishment characteristics, participation in labor market programs, and periods of unemployment. In the analyses, this linkage does not allow any conclusions to be drawn about your identity. Naturally, all provisions of the Social Code are strictly observed.

An average consent rate around 80 percent was expected *a priori*, based on comparable data linkages with administrative data from the IAB. For instance, in the most recent linkage with the Socio-Economic Panel (SOEP) data, the average consent rate for linkage with IAB data across all samples was 74 percent (Antoni et al., 2025). As shown in Table 1, the data indicate an average consent rate of around 83 percent in AID:A 2023, thus exceeding the initial expectations.

**Table 1: Requested and given consent for record linkage in AID:A 2023**

| | parents | mothers | fathers | non-parents aged 16-36 | total |
|---|---|---|---|---|---|
| Asked for consent | 6,114 | 3,432 | 2,680 | 3,006 | 9,141 |
| Consent given | 5,075 | 2,836 | 2,237 | 2,487 | 7,582 |
| Share | 83.01% | 82.63% | 83.47% | 82.73% | 82.94% |

***Source:*** **AID:A 2023 (Kuger et al., 2023), own calculations.** ***Note:*** **The calculation is exclusively based on individuals with a completed personal interview and for whom complete information is available on age and on parenthood status with respect to target persons under 18. Two parents identifying as “diverse” are included in the “parents” and “total” columns, but excluded from “mothers” and “fathers.”**

## 2.2 Possible biases in willingness to consent

We analyze whether respondents who provided consent for data linkage differ systematically in certain characteristics from respondents who did not give consent. If such selectivity exists within the subgroup that provided consent, results based on this subgroup cannot be readily generalized to the entire survey population.

To identify possible selectivity, we compare various individual characteristics between respondents with and without consent to data linkage using simple t-tests. The groups of parents (Table 2) and non-parents (Table 3) were analyzed separately.

For both groups, the results show that neither living in a large city or in East Germany nor the age of respondents is significantly associated (at the 5-percent significance level or lower) with the willingness to provide consent. Gender shows a statistically significant but relatively small association, and this only within the group of non-parents.

For both groups–parents and non-parents–there is a statistically significant correlation between willingness to consent and immigration background, survey mode, and whether respondents belong to the panel sample or to the group first surveyed in 2023, the so-called refreshment sample.

For the parent group, several additional characteristics were examined: there is no significant relationship between consent to data linkage and family type. The relationship with employment status is only weakly significant (at the 10-percent significance level) and also rather small in magnitude. A more robust relationship is observed with respect to the highest level of educational attainment.

**Table 2: Mean characteristics by willingness to consent, t-test of differences, parents**

| | (1)<br>no consent | (2)<br>consent | (3)<br>difference (1)-(2) |
|---|---|---|---|
| *Survey mode:* | | | |
| CAPI | 0.51 | 0.45 | 0.06*** |
| CATI | 0.28 | 0.37 | -0.09*** |
| CAWI | 0.15 | 0.11 | 0.05*** |
| Panel respondent | 0.41 | 0.45 | -0.04** |
| Female | 0.57 | 0.56 | 0.01 |
| Age | 42.06 | 41.55 | 0.51* |
| Born in Germany | 0.78 | 0.84 | -0.05*** |
| Urban ($\geq$ 100,000 inhabitants) | 0.64 | 0.65 | -0.01 |
| East Germany | 0.19 | 0.19 | -0.00 |
| University entrance qualification (*Abitur*) | 0.44 | 0.48 | -0.05*** |
| Employed | 0.83 | 0.85 | -0.02* |
| Couple household with biological/adopted children only | 0.86 | 0.85 | 0.01 |
| N | 1,039 | 5,075 | 6,114 |

***Source:* AID:A 2023 (Kuger et al., 2023), own calculations. *Note:* *** $p<0.01$, ** $p<0.05$, * $p<0.1$. The calculation is exclusively based on individuals with a completed personal interview and for whom complete information is available on age and on parenthood status with respect to target persons under 18.**

In summary, the factors survey mode, refresher status, immigration background, and education are significantly associated with consent to data linkage. Although these differences are only very moderate, in data analysis, these factors should be methodologically addressed through an appropriate selection correction to prevent biased inferences about the overall survey population.

**Table 3: Mean characteristics by willingness to consent, t-test of differences, non-parents aged 16-36**

| | (1) no consent | (2) consent | (3) difference (1)-(2) |
|---|---|---|---|
| *Survey mode:* | | | |
| CAPI | 0.51 | 0.37 | 0.14*** |
| CATI | 0.26 | 0.44 | -0.17*** |
| CAWI | 0.16 | 0.14 | 0.02 |
| Panel respondent | 0.38 | 0.50 | -0.13*** |
| Female | 0.43 | 0.49 | -0.05** |
| Age | 22.89 | 22.83 | 0.06 |
| Born in Germany | 0.87 | 0.91 | -0.05*** |
| Urban ($\geq$ 100,000 inhabitants) | 0.65 | 0.66 | -0.01 |
| East Germany | 0.17 | 0.19 | -0.02 |
| N | 519 | 2,487 | 3,006 |

***Source:*** **AID:A 2023 (Kuger et al., 2023), own calculations.** ***Note:*** ***** p$<$0.01, ** p$<$0.05, * p$<$0.1. The calculation is exclusively based on individuals with a completed personal interview and for whom complete information is available on age and on parenthood status with respect to target persons under 18.**

# 3 Linkage identifiers

## 3.1 Identifiers from the AID:A survey)

After the initial data preparation dealing with duplicate entries, it was possible to use the following linkage identifiers:

- First name and surname
- Month and year of birth
- Sex
- Address (zip code, place name, street name, house number)

Some respondents had more than one record within these data, which could be identified as belonging to the same person by an anonymous unique person identification number (*ID_linkage*) that was generated and assigned by infas (Institute for Applied Social Science) prior to the match.

## 3.2 Administrative data of the German Federal Employment Agency

The linkage identifiers used to find linkage consenters within the administrative data of the Federal Employment Agency (BA) were extracted from the BA's data warehouse. These data were delivered to the FDZ by IAB's department *Data and IT-Management* (DIM). These data originate from the following sources:

- Mandatory social security notifications by employers about any of their employees subject to social security contributions (i.e., not including self-employed and civil servants)
- Internal processes of the BA regarding
  - Benefit recipients according to Social Code Books II and III
  - Registered job-seekers
  - Participants in active labor market policy measures

Data entries from any of these sources related to the same person are already integrated by the statistics department of the BA, where every individual represented in the data also receives a unique and time-consistent identification number. Note that this unique pseudonym is only valid within the data of the BA and cannot be used to identify persons in any external data source. The set of linkage identifiers from the records of the BA used in this linkage was consistent to that from the survey data.

To uphold the principle of data economy and to make the linkage process more efficient, we restricted the data extract from the records of the BA. This extract only contained records of people from birth cohorts and with postcodes (coarsened to the first three digits) that actually occurred among the respondents who had provided linkage consent.

# 4 Pre-processing

Although both data sources originate from data generating processes that already include some quality checks, errors and inconsistencies during the collection process cannot be completely ruled out. Therefore, the linkage identifiers described above may still have included some errors. Using them for a comparison without any prior cleaning would thus have lead to a sub-optimal linkage result. We therefore used a set of pre-processing scripts developed by the GRLC to clean and standardize name and address fields. These routines are described in more detail by Bachbauer (2024).

The basic steps to modify string variables in both data sources included the following:

- Replacing German umlauts and other non-ASCII characters with ASCII equivalents
- Changing all characters to uppercase
- Removing leading and trailing blanks
- Removing punctuation and special characters

It was crucial to apply all steps of the pre-processing consistently on the linkage identifiers of both data sources to be linked to maximize comparability.

# 5 Comparison and classification

After the pre-processing of linkage identifiers we performed several steps to compare the record pairs from the two data sources. This section briefly describes the methods applied in these comparisons.

**Deterministic linkage:** In a deterministic linkage step, all or a predefined subset of identifiers have to be fully identical for a record pair to be classified as a match. We perform this comparison as a set of simple *merge*-operations in Stata.

**Blocking:** Blocking limits the number of necessary comparisons of record pairs and thereby considerably reduces computation time. This is achieved by only comparing record pairs that have equal values on one or a set of blocking variables. As the deterministic linkage is computationally much less demanding than the distance-based linkage, we only apply blocking in the distance-based steps.

**Distance-based linkage:** In real-world data, a considerable share of records usually still has errors of some sort even after the pre-processing, e.g., typos, misspellings, inconsistent use of abbreviations, or different orderings of name components. In a deterministic linkage setting, even the smallest deviation between identifiers of a record pair leads to a classification as a non-link, even though the record pair might in fact represent a true match. Distance-based linkage deals with such deviations by computing the similarity of different identifier representations and by matching record pairs that exceed a certain similarity threshold. In the distance-based linkage steps described below, we use Jaro-Winkler string comparator[1]

[1] This string comparator determines the similarity of a pair of strings based on how many letters they share in the same position and in the same order. The Jaro-Winkler string comparator, a variant of the Jaro´s string comparator, puts a higher weight on characters at the beginning of strings.

for identifiers containing letters and an exact comparison for numerical identifiers. We use the Merge ToolBox (MTB) software[2] to perform the distance-based linkage.

**Probabilistic linkage:** The probabilistic linkage is an extension of the distance-based linkage in which the similarity of identifiers are not simply all added up. Identifiers are weighted according to their variance within the population under consideration. This is done because different identifiers vary in how strongly a given agreement is indicative for whether a record pair might actually be a true link. For instance, an agreement on the surname indicates the likelihood of a true match more strongly than an agreement on sex. Inversely, disagreement on the identifier 'sex' indicates a non-match more strongly than disagreement of the identifier 'surname' because of the latter's higher discriminatory power. See Table 4 for an overview of which comparative metric was used for which linkage identifier, and the m- and u-parameters applied in the probabilistic linkage. We also used the MTB for this step.

**Table 4: Overview of linkage identifiers, comparative metrics and parameters**

| Linkage identifiers | comparative metrics | m-parameter | u-parameter |
|---|---|---|---|
| First name | Jaro-Winkler | 0.801 | 0.002 |
| Surname | Jaro-Winkler | 0.85 | 0.0005 |
| Birth date (day) | exact | 0.9 | 0.1 |
| Birth date (month) | exact | 0.967 | 0.08 |
| Birth date (year) | exact | 0.978 | 0.02 |
| Sex | exact | 0.988 | 0.5 |
| Zip code | exact | 0.889 | 0.049 |
| Place name | Jaro-Winkler | 0.876 | 0.012 |
| Street name | Jaro-Winkler | 0.792 | 0.001 |
| House number | exact | 0.821 | 0.02 |

**Array matching:** When an identifier has more than one possible representation within one of the compared datasets, array matching can be applied to increase linkage success. An array match compares all representations of a given identifier for a record in one dataset with all representations of that identifier in a record in the other dataset and only considers the highest similarity value as the result of the comparison. The function of array matching is not commonly available in record linkage software, but it is one of the options available in the MTB.

2 The MTB is maintained by the GRLC and can be downloaded and used for free for academic purposes. See https://record-linkage.de or Schnell/Bachteler/Bender (2004) for more details on the software.

# 6 Linkage results

## 6.1 Linkage success rates

Table 5 summarizes the linkage steps we performed consecutively, starting with the strictest criterion of an exact agreement on all available identifiers (step *Deterministic 1*). This first step links persons based on the exact agreement of the first name, surname, sex, birth date (month, year), address (zip code, place name, street name, house number). In this step, 81.1 percent of the 7,593 persons available in the data extract from infas are already linked successfully. The exact matching was then repeated iteratively by excluding one of the following (sets of) identifiers: zip code, place name, street name, house number, year of birth & month of birth. These six sub-steps are summarized in the step *Deterministic 2*. Relative to the first step, the cumulative contribution of these steps is rather small, as they only add 6.1 percentage points to the overall linkage rate.

Only persons not successfully linked in the deterministic steps are compared in the distance-based steps. To reduce computation time, blocking is used in all distance-based steps. *Distance-based 1*, which uses the 5-digit zip code as a blocking variable, adds 7.21 percentage points to the linkage rate. *Distance-based 2* uses blocking on month and year of the birth date as well as on sex, and it achieves additional 0.59 percentage points. Using the NYSIIS (New York State Intelligence and Identification System) phonetic code[3] on the first name as the blocking variable in *Distance-based 3* added another 1.1 percentage points. The step *Distance-based 4* using blocking on the 3-digit zip code added 0.34 percentage points to the linkage rate. *Distance-based 5* used blocking on the year of birth, and only added 0.04 percentage points to the linkage rate. Record pairs that could only be classified as potential matches throughout the distance-based steps were collected for the *Manual classification*, which finally added another 3.47 percentage points to the overall linkage rate.

Of the 7,593 respondents that could have been linked, we were able to find matches for 7,268 (95.72 percent), which is a very high success rate compared to many other linkages with IAB data.[4] However, Table 6 shows that not all of these matches are available in the linked research data. For 357 of the matched cases, the Integrated Employment Biographies (IEB) of the IAB did not contain any records even though the Data Warehouse of the BA[5] did provide

3 NYSIIS is a phonetic code used to encode character strings such as names. See Taft (1970) for more details.
4 For example, Dummert/Sauer (2024) and Bachbauer (2024) report linkage success rates of 85.9 percent and 88.4 percent, respectively.
5 We differentiate between the data from Data Warehouse of the BA, which contains the linkage identifiers originating from the sources listed in Section 3.2, and the research data in the IEB. The latter is created by IAB's department DIM based on extracts from the BA's Data Warehouse, but it only contains pseudonomized

**Table 5: Summary of linkage steps**

| Linkage steps | N | share | description |
|---|---|---|---|
| | 7,593 | 100.00% | Number of persons in AID:A data before linkage |
| Deterministic 1 | 5,896 | 81.12% | Exact agreement on first name, surname, sex, birth date (month, year), address (zip code, place name, street name, house number) |
| Deterministic 2 | 445 | 6.12% | Exact agreement on identifiers as in *Deterministic 1* while excluding single identifiers in each of the six separate sub-steps (zip code; place name; street name; house number; year of birth & month of birth) |
| Distance-based 1 | 524 | 7.21% | Comparison of first name, surname, sex, birth date (month, year), address (zip code, place name, street name, house number); blocking on 5-digit zip code |
| Distance-based 2 | 43 | 0.59% | Comparison of first name, surname, address (zip code, place name, street name, house number); blocking on year and month of birth and sex |
| Distance-based 3 | 80 | 1.10% | Comparison of first name, surname, sex, birth date (month, year), address (zip code, place name, street name, house number); blocking on NYSIIS-code of first name |
| Distance-based 4 | 25 | 0.34% | Comparison of first name, surname, sex, birth date (month, year), address (zip code, place name, street name, house number); blocking on 3-digit zip code |
| Distance-based 5 | 3 | 0.04% | Comparison of first name, surname, sex, birth date (month), address (zip code, place name, street name, house number); blocking on year of birth |
| Manual classification | 252 | 3.47% | Clerical review of all remaining potential matches and manual classification of true matches. |
| Total | 7,268 | 95.72% | |

***Source:* Data warehouse of the BA, infas data; own calculations. *Note:* All distance-based steps use probabilistic linkage with the comparative metrics and parameters shown in Table 4.**

linkage identifiers for the underlying persons. This may be the case when a person has contact with a local employment agency (e.g., for career counseling), during which their address gets recorded, but they have subsequently not (yet) entered any of the labor market states that are relevant for the IEB.

**Table 6: Linkage success and availability of research data**

| | N | share |
|---|---|---|
| Linkage candidates | 7,593 | 100.00% |
| Successfully linked | 7,268 | 95.72% |
| Successfully linked, IAB research data available | 6,911 | 91.02% |

***Source:* Data warehouse of the BA, infas data; own calculations.**

research data. See Schmucker/vom Berge (2025) for more details on these research data.

## 6.2 Possible biases in linkage success

Among respondents with linkage consent, we now explore possible biases in the course of the linkage process. That is, we compare various individual characteristics between consenting respondents whose survey data could and could not be successfully matched with administrative data. Again, we use simple t-tests and analyze the groups of parents (Table 7) and non-parents (Table 8) separately.

Concerning the group of parents, the patterns are surprisingly similar to those found for linkage consent bias in Table 2: differences between those who were successfully matched and those who were not are mainly observed regarding survey mode, refresher-status, immigration status, and education (however, in the reverse direction). Age plays a more prominent role in linkage success bias than it does in consent bias, where differences are only weakly statistically significant.

For the group of non-parents, there are differences in survey mode between unmatched and matched respondents, similar to what has been found concerning consent bias (see Table 3). There are no similarities with respect to the remaining characteristics. The only other relevant differences are observed for respondent age. Age also played a role for the group of parents, however, in the reverse direction: while unmatched parents are on average older than matched parents, among non-parents we find the unmatched to be younger.

**Table 7: Mean characteristics by match status, t-test of differences, parents**

| | (1) no match | (2) match | (3) difference (1)-(2) |
|---|---|---|---|
| *Survey mode:* | | | |
| CAPI | 0.51 | 0.44 | 0.07* |
| CATI | 0.28 | 0.38 | -0.10*** |
| CAWI | 0.14 | 0.11 | 0.04 |
| Panel respondent | 0.37 | 0.46 | -0.09** |
| Female | 0.54 | 0.56 | -0.02 |
| Age | 43.50 | 41.46 | 2.04*** |
| Born in Germany | 0.68 | 0.84 | -0.16*** |
| Urban ($\geq$ 100,000 inhabitants) | 0.66 | 0.65 | 0.01 |
| East | 0.16 | 0.20 | -0.04 |
| University entrance qualification (*Abitur*) | 0.56 | 0.48 | 0.08** |
| Employed | 0.80 | 0.85 | -0.05* |
| Couple household with biological/adopted children only | 0.84 | 0.86 | -0.01 |
| N | 221 | 4,854 | 5,075 |

***Source:* AID:A-ADIAB 7523, own calculations. *Note:* *** p$<$0.01, ** p$<$0.05, * p$<$0.1. The calculation is exclusively based on individuals with linkage consent, with a completed personal interview, and for whom complete information is available on age and on parenthood status with respect to target persons under 18.**

**Table 8: Mean characteristics by match status, t-test of differences, non-parents aged 16-36**

| | (1) no match | (2) match | (3) difference (1)-(2) |
|---|---|---|---|
| *Survey mode:* | | | |
| CAPI | 0.31 | 0.38 | -0.07*** |
| CATI | 0.42 | 0.44 | -0.02 |
| CAWI | 0.19 | 0.13 | 0.07*** |
| Panel respondent | 0.53 | 0.50 | 0.03 |
| Female | 0.50 | 0.48 | 0.01 |
| Age | 18.40 | 23.84 | -5.44*** |
| Born in Germany | 0.92 | 0.91 | 0.01 |
| Urban ($\geq$ 100,000 inhabitants) | 0.68 | 0.66 | 0.02 |
| East Germany | 0.21 | 0.18 | 0.02 |
| N | 462 | 2,025 | 2,487 |

***Source:* AID:A-ADIAB 7523, own calculations. *Note:* *** p$<$0.01, ** p$<$0.05, * p$<$0.1. The calculation is exclusively based on individuals with linkage consent, with a completed personal interview, and for whom complete information is available on age and on parenthood status with respect to target persons under 18.**

# 7 Discussion

The linkage consent rate among AID:A respondents was 82.9 percent which is high in comparison to comparable survey-linkages with IAB data. The overall linkage success rate was about 95.7 percent, which is above the average of previous linkage applications with data of the IAB. Given that the administrative data of the BA only cover roughly 85 percent of the German workforce, the linkage can be classified as very successful.

Explorations of linkage consent bias indicate some degree of selectivity in respondents' willingness to consent, which should be taken into account in subsequent analyses. However, the observed differences are relatively small in magnitude. Analyses of linkage success bias suggest that mainly age differences between matched and non-matched respondents are relevant for further analyses, particularly when it comes to the generalizability of results.

# References


Antoni, Manfred; Beckmannshagen, Mattis; Grabka, Markus M.; Keita, Sekou; Trübswetter, Parvati (2025): Befragungsdaten der SOEP-Core-, IAB-SOEP Migrationsstichprobe, IAB-BAMF-SOEP Befragung von Geflüchteten und SOEP-Innovationssample verknüpft mit administrativen Daten des IAB (SOEP-CMI-ADIAB) 1975-2021. FDZ-Datenreport 01/2025 (de), DOI: 10.5164/IAB.FDZD.2501.de.v1.

Antoni, Manfred; Schnell, Rainer (2019): The Past, Present and Future of the German Record Linkage Center (GRLC). In: Journal of Economics and Statistics, Vol. 239, No. 2, p. 319–331, DOI: 10.1515/jbnst-2017-1004.

Bachbauer, Nadine (2024): NEPS-SC3 survey data linked to administrative data of the IAB (NEPS-SC3-ADIAB). FDZ-Datenreport 02/2024 (en), DOI: 10.5164/IAB.FDZD.2402.en.v1.

Christen, Peter (2012): Data Matching: Concepts and Techniques for Record Linkage, Entity Resolution, and Duplicate Detection. Berlin: Springer, DOI: 10.1007/978-3-642-31164-2.

Dummert, Sandra; Sauer, Irakli (2024): PASS-Befragungsdaten verknüpft mit administrativen Daten des IAB (PASS-ADIAB) 1975-2022. FDZ-Datenreport 05/2024 (de), DOI: 10.5164/IAB.FDZD.2405.de.v1.

Kuger, Susanne; Bernhardt, Janine; Herz, Andreas; Linberg, Anja; Recksiedler, Claudia; Simm, Inga; Ulrich, Susanne; Braun, Dominik; Blindow, Katrina; Fischer, Andreas; Schacht, Diana; Quellenberg, Holger; Berngruber, Anne; Chabursky, Sophia; Eichhorn, Thomas; Entleitner-Phleps, Christine; Guglhoer-Rudan, Angelika; Hofherr, Stefan; Langmeyer, Alexandra; Maly-Motta, Hanna; Naab, Thorsten; Schüller, Simone; Steinberg, Hannah; Steiner, Christine; Tran, Kien; Zerle-Elsäßer, Claudia; Walper, Sabine (2023): Aufwachsen in Deutschland: Alltagswelten 2023 (AID:A 2023). Haupterhebung, 2023. Onsite Version: 1.0.0. Dataset. DOI: 10.17621/aida2023.

Schmucker, Alexandra; vom Berge, Philipp (2025): Sample of Integrated Labour Market Biographies (SIAB) 1975-2023 v2. FDZ-Datenreport 02/2025 v2 (en), DOI: 10.5164/IAB.FDZD.2502.en.v2.

Schnell, Rainer; Bachteler, Tobias; Bender, Stefan (2004): A toolbox for record linkage. In: Austrian Journal of Statistics, Vol. 33, p. 125–133.

Taft, Robert L (1970): Name search techniques. Bureau of Systems Development, New York State Identification and Intelligence System.

# List of Tables

# Imprint

**FDZ-Methodenreport 07|2025 EN**

**Date of publication**
31 October 2025

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



**Download**
https://doku.iab.de/fdz/reporte/2025/MR_07-25_EN.pdf

**DOI**
10.5164/IAB.FDZM.2507.en.v1

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

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

---

**Corresponding author**
Manfred Antoni
Email: manfred.antoni@iab.de