

Data Observer

Manfred Antoni* and Arne Bethmann

PASS-ADIAB – Linked Survey and Administrative Data for Research on Unemployment and Poverty

<https://doi.org/10.1515/jbnst-2018-0002>

1 Introduction

We present a dataset called “PASS-ADIAB - PASS survey data linked to administrative data of the IAB”¹ which integrates survey data from the household panel study “Labour Market and Social Security” (PASS, see Trappmann et al., 2013) with administrative labour market biographies from registers at the German Federal Employment Agency (BA) using record linkage techniques. The dataset is provided by the Research Data Centre (FDZ) of the Federal Employment Agency at the Institute for Employment Research (IAB) and is available for on-site use as well as through remote data access.

Combining the two data sources enhances research opportunities for a wide range of socioeconomic topics, especially in the fields of unemployment and poverty research. PASS provides ten waves of survey data from household and individual interviews on a wide variety of issues relating to the socioeconomic situation. Administrative biographies complement these data with the full history of individual labour market records concerning employment, unemployment, programme participation etc.

¹ See http://fdz.iab.de/en/FDZ_Individual_Data/PASS.aspx and Antoni et al. (2017) for further details.

***Corresponding author: Manfred Antoni**, Research Data Centre (FDZ), Institute for Employment Research (IAB), Regensburger Str. 104, 90478 Nuremberg, Germany, E-mail: manfred.antoni@iab.de

Arne Bethmann, Munich Center for the Economics of Aging (MEA), Max Planck Institute for Social Law and Social Policy, Amalienstr. 33, 80799 Munich, Germany, E-mail: a.bethmann@mea.mpisoc.mpg.de

2 Research opportunities

PASS-ADIAB offers a unique combination of high quality, long running household panel survey data on poverty, unemployment, and welfare with detailed data on individual labour market biographies from administrative registers. While administrative data can give exact accounts of welfare dynamics themselves, augmenting them with survey data can foster insights concerning the impact of more subjective measures (e.g. reservation wages) on these processes. The same is true for welfare dynamics with regard to household context. Apart from structural information like household composition, the survey data contain detailed accounts of the household situation, e.g. on material deprivation and child care arrangements. On the other hand enriching survey data with full administrative biographies enhances the validity of research into, for instance, the interplay of health and unemployment. Another important benefit of the administrative data is reliable longitudinal information on the earnings and benefit receipt of linked respondents.

The data are also well suited for a wide range of research questions going beyond aspects of welfare and poverty dynamics. The linked PASS surveys include detailed information on individual and more subjective topics not found in the register data, including:²

- Attitudes (e.g. towards employment, minimal wages, managing finances, gender roles, reciprocity)
- Job quality
- Interaction with job-centers
- Life satisfaction (general and domain specific) and social inclusion
- Social networks
- Health issues (e.g. SF-12 Physical and mental health scale)
- Psychological traits (e.g. Big Five, self-efficacy, impulsiveness)
- Migration background

In addition to substantive research, PASS-ADIAB provides ample opportunity for methodological research. Several measures are collected in both the survey and the administrative data. This allows for analyses of, for instance, underreporting of welfare receipt (see Bruckmeier et al. 2014; Kreuter et al. 2010). Selection processes in the linkage procedure, which might influence the validity of results, can

² The full survey instruments for all PASS surveys are available here: http://fdz.iab.de/en/FDZ_Individual_Data/PASS.aspx

be analysed using characteristics of survey respondents (see Beste 2011). Interviewer effects can be evaluated via an anonymised interviewer identifier in the survey data.

3 Data sources

3.1 PASS household panel survey

In 2005 German labour market reforms – especially the introduction of Unemployment Benefit II (UB II) – prompted an additional demand for research into the effects of welfare policies. PASS was established at IAB shortly thereafter funded by the German Federal Ministry of Labour and Social Affairs in order to provide a new data source for welfare research. The study's main focus is on topics related to pathways into and out of UB II receipt, the dynamics of the material and social situation of individuals and households as well as behaviour and attitudes of individuals over time.³

PASS is an annual panel study that has been conducted by the IAB since 2006. Interviews are conducted on the household level as well as with every individual in the household aged 15 or older. Data collection is done using a sequential mixed mode design starting in a face to face mode (CAPI) and followed up by telephone (CATI) if the household or individual could not be contacted or prefers a telephone interview. Due to a large number of individuals with migration background in the sample, Turkish and Russian translations of the questionnaire were used if necessary.

The initial survey sample consisted of two subsamples. The first one is a sample of households containing at least one welfare benefit recipient drawn from the UB II register at the Federal Employment Agency. The other one was drawn from the German resident population. This sampling scheme gives higher statistical power in analysing welfare recipients but still allows for projections to the German resident population when using the proper weights included in the dataset. In all consecutive waves a refreshment sample was drawn from the UB II population in order to correct for changes in the composition of UB II recipients. In wave 5 (2011) an additional refreshment sample was drawn for both populations. The sampling was conducted via a two stage approach with postal codes as the primary sampling unit and sampling probabilities proportional to size (PPS).

While the survey data provide information on a wide variety of topics relevant to unemployment, poverty, and welfare research, most of these data are

³ For more details on the PASS study see Trappmann et al. (2013) and Bethmann et al. (2016).

subjective in nature since they were reported by respondents and are therefore prone to errors arising from the response process (see Tourangeau et al. 2000). This might be especially problematic for topics that individuals are reluctant to report on as is the case for welfare receipt (see e.g., Bruckmeier et al. 2015). But even if respondents are willing to give proper answers, retrospective reporting of labour market histories is likely to produce errors, even more so as individuals get older and biographies increase in length and complexity. This is where administrative data can add considerably to data quality and hence improve the validity of substantive conclusions drawn from analyses.

3.2 Administrative labour market biographies

The PASS survey data are augmented by administrative data available at the IAB. These data contain every person in Germany that, at some point since 1975, held one of the following employment statuses: employment subject to social security (since 1975), marginal part-time employment (since 1999), benefit receipt according to the German Social Code III (since 1975) or II (since 2005), registered job-seeking (since 1997) and (planned) participation in programmes of active labour market policies (since 2000).

The administrative data have two sources: first, employment spells stem from compulsory notifications by employers to the social security system. Such notifications on their employees have to be given by each employer at least once a year or when the employment relationship ends before the end of the year. Notifications in the course of the year have to be given when any of the characteristics required by the notification scheme change (e.g., the health insurance company or residential address of a given employee). Second, the data on benefit receipt, job search and participation in labour market programmes are mainly entered by the caseworkers at the local employment agencies while registering these statuses or providing the corresponding services. Observations from each of these sources have a longitudinal structure with start- and end-dates that are detailed to the day.

Employment spells can be enriched with administrative establishment data.⁴ These data contain detailed and reliable information on establishment characteristics, the structure of employees (e.g., by age, sex, qualification, occupation, type of employment) and the average wage of full-time employees. These characteristics are given for establishments at which the jobs included in the employee

⁴ The structure of the establishment data contained in PASS-ADIAB is identical to that of the Establishment History Panel (BHP). For more details see Schmucker et al. (2016) or http://fdz.iab.de/en/FDZ_Establishment_Data/Establishment_History_Panel.aspx.

data are held. Like the employee data, the establishment data have a longitudinal structure. Instead of daily spell data, the establishment data are created as yearly cross-sections with a reference-date of June 30th.

In addition to these yearly cross-sectional files, additional extension files may be requested. On the one hand, they contain the structure of inflows and outflows of employees (e.g., by sex, type of employment, occupational groups, age). On the other hand, the extension files contain information on entries and exits of the establishments themselves. These variables allow a distinction between openings and closings of establishments from, for instance, changes of establishment numbers merely due to restructuring or relabeling of the establishments' superordinate firms.

Generally, these administrative data are very reliable, especially for information that is directly relevant for the amount of unemployment or pension insurance claims. However, some characteristics are only reported for statistical purposes in some sources, while being essential for other sources. For example, information regarding a person's qualification or occupation is highly important for appropriate job offers sent out by caseworkers. In contrast, these characteristics have only statistical value when reported for employment relationships. Many employers therefore put in less effort in keeping such variables up to date compared to when reporting on, for instance, a person's wage sum during the year. The latter, on the other hand, is directly relevant for social security contributions and unemployment insurance entitlements. This leads to a somewhat higher share of cases with missing or potentially outdated data on some characteristics. Variables that are subject to these potential shortcomings are documented in great detail by Antoni et al. (2016) and Schmucker et al. (2016). Figure 1 provides an overview of the data sources of PASS-ADIAB.

4 Record linkage

The starting point of the record linkage (Christen 2012) are all participants of the PASS survey that provided informed consent to the linkage of their survey data with information given on them within the administrative data of the IAB. The question for linkage consent was asked in every wave of the PASS survey since its beginning, with an average consent rate over waves 1–8 of 81%.

The linkage of new consenters from waves 1–5 was performed in separate rounds after each of these waves had been concluded, whereas the linkage for waves 6–8 was done jointly after wave 8. During that time, alternating people and departments were responsible for different steps of the linkage process. Besides the department responsible for the PASS study and the FDZ, the main contributors

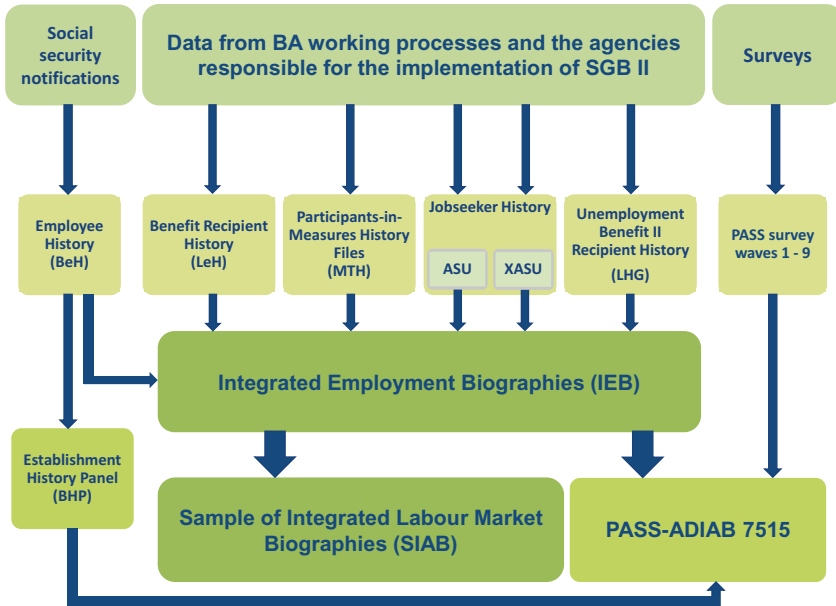


Figure 1: Data sources of PASS-ADIAB.

Source: translated from Antoni et al. (2017).

were staff of the IAB department Data and IT-Management and the German Record Linkage Center (GRLC, see Antoni and Schnell 2017). The GRLC was responsible for the linkage rounds of waves 1 and 5–8.

The methods of linkage, their combination and their sequence vary between the two types of samples and the linkage rounds. During each of the linkage rounds, a so-called Goldstandard linkage was attempted for all consenters of the UB II sample. This linkage takes advantage of the fact that the UB II sample was drawn from the administrative data in the first place. For these cases, the linkage only had to identify members within a given benefit community (Bedarfs-gemeinschaft), which only requires the name, sex and the birth date of a person. For the remaining consenters, deterministic and probabilistic methods were used to compare the name, sex, birth date and address of consenters and to identify them in the IAB's administrative data. Antoni et al. (2017) provide more details on the applied methods and the overall success rate of all linkage rounds.

For the result of the first linkage round, Beste (2011) examines potential linkage consent bias and whether linkage success is selective. His results show very low consent bias, but they hint at some selectivity in linkage success regarding age, educational level or employment status of respondents. Nevertheless,

in further analyses he verifies that the existing bias does not influence research results. To do so, he estimates exemplary models, once based on all respondents and alternatively only based on successfully linked respondents. A comparison of these results shows no significant differences in the findings based on the different estimation samples.

5 Data structure

PASS-ADIAB consists of the PASS Scientific Use File (SUF, see Bethmann et al. 2016) and several datasets containing the administrative labour market biographies for all PASS respondents that could be linked. Table 1 gives an overview of the number of household and individual interviews in each of the nine panel waves as well as the number of individual interviews with linked administrative data available.⁵

It is important to bear in mind that, although only consenters of waves 1 through 8 have been linked so far, PASS-ADIAB also contains survey data from PASS wave 9. The numbers of individuals and households in this part of PASS-ADIAB is therefore identical to those in the PASS SUF released after wave 9. While the survey data in PASS-ADIAB partially cover the year 2015, the administrative data only cover the years 1975–2014.

The two main datasets in the PASS SUF are the one containing the household interviews (HHENDDAT) and the one containing the individual interviews (PENDDAT). These are both provided in long format, i.e. each row in the dataset

Table 1: Numbers of observations in PASS-ADIAB by wave.

	Households	Individuals	Linked individuals
Wave 1	12794	18954	13514
Wave 2	8289	12487	9294
Wave 3	9345	13439	10069
Wave 4	7739	11768	8975
Wave 5	10185	15607	11132
Wave 6	9455	14619	10556
Wave 7	9460	14449	10575
Wave 8	8946	13460	9790
Wave 9	8878	13271	8553
Total	24047	37993	25048

Source: PASS-ADIAB 7515, own calculation.

⁵ Only individual interviews can be linked to the administrative data.

contains the responses collected from an individual (or household) in a specific wave. Individuals and household can be identified by their unique identifiers (*pnr* and *hnr*). Apart from the two main datasets the PASS SUF includes register datasets, datasets with weights, self-reported welfare and labour market biographies (only for a two year retrospective period) and a few others. The PASS User Guide (Bethmann et al. 2013) gives in-depth information on the separate datasets comprising the SUF and on how to work with them.

The administrative data in PASS-ADIAB, whose structure is identical to that of the Sample of Integrated Labour Market Biographies (SIAB, see Antoni et al. 2016), consist of a number of files. The main file is the *individual file* that contains the longitudinal labour market biographies of all linked respondents. In order to use the linked administrative data they have to be merged to the individual interviews from the PASS SUF prior to analysis. To do so, one merges records that belong to the same person identifier *pnr*, which is contained in the survey datasets as well as in the administrative *individual file*. This is also shown in the User Guide: the procedure applicable to administrative data is similar to using PASS's own biography dataset (Stata: *ibid*, Example 9.4; SPSS: Fuchs et al. 2015, Example 1.4).

For each employment record in the *individual file*, basic yearly information on the establishment a respondent was employed at is stored in the *establishment file*. Several extension files contain additional variables on the establishment level as well as on worker flows (inflows/outflows) and establishments dynamics (entries/exits). All establishment variables can be linked to employment records in the *individual file* using the establishment identifier *betnr*.

Using the person identifier *pnr*, researchers can add variables on technical aspects of the linkage procedure on the individual level. These variables as well as the person identifier *pnr* are stored in a separate *linkage quality file*. Figure 2 gives an overview of the data structure of PASS-ADIAB.

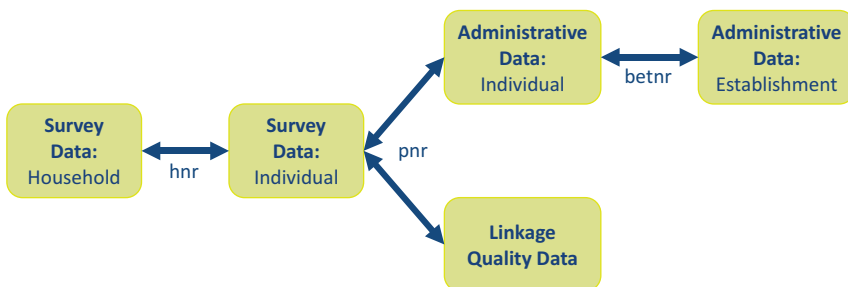


Figure 2: Data structure of PASS-ADIAB.

Source: own illustration.

6 Data access

Due to the comprehensive information on linked respondents provided in the linked data, there is no Scientific Use File of PASS-ADIAB. To assure privacy for respondents on the one hand and to retain the analytic potential of the linked data on the other hand, PASS-ADIAB can only be used via access modes that provide the highest level of data security. PASS-ADIAB therefore is available to researchers via on-site use at one of FDZ's locations both in Germany and abroad⁶ and via subsequent remote data access using the application JoSuA (see Eberle et al. 2017).

To achieve data access, researchers have to submit a request form to the FDZ via email. Any requests are verified for compliance with the provisions of Section 75 of the German Social Code Book X by the FDZ and need approval by the Federal Ministry of Labour and Social Affairs. The whole procedure usually takes up to one month.⁷

References

- Antoni, M., S. Dummert, S. Trenkle (2017), PASS-Befragungsdaten verknüpft mit administrativen Daten des IAB (PASS-ADIAB) 1975–2015. FDZ Datenreport 06/2017 (de).
- Antoni, M., A. Ganzer, P. vom Berge (2016), Sample of Integrated Labour Market Biographies (SIAB) 1975–2014. FDZ Datenreport 04/2016 (en).
- Antoni, M., R. Schnell (2017), The Past, Present and Future of the German Record Linkage Center (GRLC). *Journal of Economics and Statistics*, online first.
- Beste, J. (2011), Selektivitätsprozesse bei der Verknüpfung von Befragungs- mit Prozessdaten. Record Linkage mit Daten des Panels „Arbeitsmarkt und soziale Sicherung“ und administrativen Daten der Bundesagentur für Arbeit. FDZ Methodenreport 09/2011 (de).
- Bethmann, A., B. Fuchs, M. Huber, M. Trappmann, A. Reindl, M. Berg, R. Cramer, C. Dickmann, R. Gilberg, B. Jesske, M. Kleudgen (2016), Codebook and Documentation of the Panel Study 'Labour Market and Social Security' (PASS): Datenreport wave 9. FDZ Datenreport 07/2016 (en).
- Bethmann, A., B. Fuchs, A. Wurdack (2013), User Guide "Panel Study Labour Market and Social Security" (PASS): Wave 6. FDZ Datenreport 07/2013 (en).
- Bruckmeier, K., G. Müller, R. T. Riphahn (2015), Survey Misreporting of Welfare Receipt—Respondent, Interviewer, and Interview Characteristics. *Economics Letters* 129: 103–107.

⁶ See http://fdz.iab.de/en/FDZ_Data_Access/FDZ_On-Site_Use/Standorte.aspx for details on these locations.

⁷ See http://fdz.iab.de/en/FDZ_Data_Access/FDZ_On-Site_Use.aspx for more details on the application procedure.

- Bruckmeier, K., G. Müller, R. T. Riphahn (2014), Who Misreports Welfare Receipt in Surveys? *Applied Economics Letters* 21: 812–816.
- Christen, P. (2012). *Data Matching: Concepts and Techniques for Record Linkage, Entity Resolution, and Duplicate Detection*. Springer, Berlin.
- Eberle, J., D. Müller, J. Heining (2017), A Modern Job Submission Application to Access IAB's Confidential Administrative and Survey Research Data. FDZ Methodenreport 01/2017 (en).
- Fuchs, B., S. Lödel, M. Otto (2015), Quick Start File for the Panel "Labour market and social security" (PASS): Analysing the PASS Data Using SPSS/PASW. FDZ Methodenreport 08/2015 (en).
- Kreuter, F., G. Müller, M. Trappmann (2010), [Nonresponse and Measurement Error in Employment Research: Making Use of Administrative Data](#). *Public Opinion Quarterly* 74 (5): 880–906.
- Schmucker, A., S. Seth, J. Ludsteck, J. Eberle, A. Ganzer (2016), Establishment History Panel 1975–2014. FDZ Datenreport 03/2016 (en).
- Tourangeau, R., L. J. Rips, K. Rasinski (2000), *The Psychology of Survey Response*. Cambridge, Cambridge University Press.
- Trappmann, M., J. Beste, A. Bethmann, G. Müller (2013), The PASS Panel Survey after Six Waves. *Journal for Labour Market Research* 46 (4): 275–281.