Research Data Centre (FDZ)
of the German Federal
Employment Agency (BA)
at the Institute for
Employment Research (IAB)

FDZ-Methodenreport 02/2018 EN

Methodological aspects of labour market data

Providing a consistent measure of further training for the IAB Establishment Panel

Tina Hinz,
Jens Stegmaier

Bundesagentur für Arbeit

Providing a consistent measure of further training for the IAB Establishment Panel

Tina Hinz (Friedrich-Alexander University Erlangen-Nürnberg), Jens Stegmaier (IAB)

Documentation version: DOI: 10.5164/IAB.FDZM.1802.en.v1

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

Figures

Tables

Abstract

In the IAB Establishment Panel, further training activities could be reported as the number of trained workers or the number of trainings provided. About 18 % of the surveyed establishments report the number of trainings. As these establishments differ in observable characteristics, exclusion of such establishments might bias results. This documentation provides an easy and effective tool to infer the number of trained workers from the number of trainings by building on an approach by Düll and Bellmann (1998). In order to make the method accessible for other researchers we discuss possibilities and limitations of this method and provide the stata code.

Zusammenfassung

Im IAB Betriebspanel können Betriebe entweder die Anzahl der weitergebildeten Beschäftigten oder die Anzahl der Weiterbildungskurse angeben. Letzteres tun etwa 18 % der befragten Betriebe. Da sich diese Betriebe in verschiedenen Merkmalen von den übrigen Betrieben unterscheiden, kann der Ausschluss dieser Betriebe zu verzerrten Ergebnissen führen. Dieser FDZ-Methodenreport stellt einen leicht umsetzbaren und effektiven Ansatz (zurückgehend auf Düll und Bellmann 1998) vor, mit dessen Hilfe die Angaben zur Anzahl der Weiterbildungskurse in Angaben zur Zahl der weitergebildeten Beschäftigten umgerechnet werden können. Wir diskutieren Möglichkeiten und Grenzen des Ansatzes und stellen ein Stata do-file zur Umrechnung zur Verfügung.

Keywords: IAB Establishment Panel, further training

Notes and thanks

The individual authors are responsible for the content of the FDZ Datenreports and Methodenreports. For valuable comments and suggestions we thank Matthias Umkehrer. We thank Hannes Walz for research assistance.

1 Introduction

The IAB Establishment Panel is a yearly panel survey of roughly 15,000 to 16,000 establishments located in Germany and is one of the most important data sources to examine employer-provided training activities. Reported information on training activities include the number of trained workers, the type of training, the qualification of the employees trained and the funding source of the training, among others.

In the IAB Establishment Panel employers usually indicate how many workers receive training. This is a valuable piece of information as it indicates the training intensity of an establishment. However, in order to minimize item non-response, employers were allowed to report the number of trainings provided instead of the number of trained workers. While most respondents report the number of workers, 18 percent of the surveyed establishments report the number of trainings. As each worker may participate in one or several training courses, it is difficult to infer how many workers were trained and there is no common way of dealing with establishments that report the number of trainings. In empirical work, some researchers ignore the observations referring to the number of trainings, they do not mention their solution in dealing with these cases or they control for the reporting mode.[1] Ignoring establishments that refer to the number of trainings may lead to biased results as these establishments differ systematically from establishments that report the number of trained workers (see the discussion in Section 2.2). Furthermore, including the establishments that report the number of trainings enlarges the size of the analysis sample. The associated gain in precision might be important for answering specific research questions.

To deal with this issue we propose to rely on a technique suggested by Düll and Bellmann (1998) for estimating the number of trained workers for establishments that report only the number of trainings. This idea has already been used by some studies (e.g. Bellmann et al. 2017, Hinz 2016, Stegmaier 2012). However, none of these studies provides an in-depth explanation of the method itself and related pitfalls.

The idea by Düll and Bellmann (1998) is, first, to cluster the establishments within relatively similar groups of establishments in terms of observable characteristics, i.e. industry, establishment size, region and year. Second, one calculates the ratio between the share of trained workers and the share of training courses within each group. Finally, one uses this ratio as a weighting factor to estimate the number of trained workers for establishments reporting the number of trainings.[2] In order to make the method accessible for other researchers we provide the stata code and discuss possibilities and limitations of this method.

This Methodenreport is structured as follows: The next section introduces the IAB Establishment Panel and the relevant training information of the survey. Section 3 discusses the approach by Düll and Bellmann (1998) and its limitations. Sections 4 and 5 provide our

[1] See e.g. Berg et al. (2017), Görlitz/Stiebale (2011), Heywood/Jirjahn/Pfister (2017) and Zwick (2006).

[2] A regression-based imputation technique would be superior to such an approach but would require establishments reporting both information at least at one point in time. This information is not available in the IAB Establishment Panel.

implementations and helpful guidelines. The final section outlines the structure of the do-file and provides operating instructions.

2 IAB Establishment Panel and the training information

2.1 IAB Establishment Panel

The IAB Establishment Panel is a large annual survey of establishments located in Germany. The survey information is collected by professional interviewers of *Kantar Public Deutschland*, who mostly visit establishments for face-to-face interviews. The survey is conducted in the third quarter of the year and a unique establishment identifier allows tracking the units of observation over time as long as establishments exist and continue to participate in the survey. The survey aims for a representative sample of about 15,000 to 16,000 establishments each year. The panel retention rate for establishments that took part in the previous year is stable at approximately 84 % for the face-to-face mode (and roughly 68 % for other modes, mainly self-completed/mail interviews) whereas only 28 % of all newly contacted establishments answer the questionnaire in face-to-face interviews (and 12 % in mail mode). The units of observation are establishments defined as workplaces that can be ascribed to a particular address. The unique establishment identifier stems from compulsory employment notifications each employer has to report for social security reasons. The IAB Establishment Panel therefore covers all establishments in Germany with at least one employee subject to social security. Establishments that have exclusively workers in marginal part-time employment are excluded from the sampling frame. For more information regarding the overall design we refer to Bechmann et al. (2017), Bossler et al. (2017) Ellguth et al. (2014) or Fischer et al. (2009).

2.2 Training information in the IAB Establishment Panel and selective reporting

Questions on further training are part of the regular program of the survey from the beginning in 1993. However, over time the structure and frequency of these questions were modified in several ways. We therefore provide a brief overview on the availability of training information (see also Table 1).

Between 1993 and 1995 the survey comprised questions on training on a yearly basis but the wording was different than in later waves. From 1997 to 2006 establishments were surveyed biennially on questions on training activities. From 2007 those questions are included annually.

The training section starts with a dichotomous question to identify establishments that trained their employees. Here, we refer to the 2013 wording:

“Did your establishment/office support training courses in the first half of 2013? To be more precise, did you release staff for the purpose of participating in internal or external training courses and did your establishment cover the expense for these in full or at least in part?”

Establishments are filtered into the section on training issues if they answered:

"Yes, staff was released and expenses were covered."

With regard to the number of trained workers a few changes were made to the questionnaire during the course of the years. From 1993 to 1995 respondents could indicate only the number of trained workers. From 1997 to 2013 they could choose to report the number of trained workers or the number of trainings. Starting with the survey wave 2014, respondents again could report only the number of trained workers. The related questions are (we refer again to the 2013 wording):

"With respect to further training courses in the first half of 2013, are you able to provide information about the number of individuals participating in the courses or rather about the number of cases of participation?"

> *"Information on individuals is given if an employee who participated in several training courses is counted several times"*
>
> *"Information on cases of participation is given if an employee who participated in several training courses is counted only once."*

"How many participants (individuals) or cases of participation were there for further training courses according to question 57 in the first half of 2013?"

It should be highlighted that the further training information in the survey always refers to the first half of a given year. This is important when relating the number of trained workers to the overall number of workers of the establishment (i.e. to calculate the intensity of trained workers), which always refers to June 30th each year. Thus, if the establishment has some turnover, it may happen that an establishment reports more trained workers than the current stock of employment on the reference date. We will come back to this issue below when we discuss the limitations of the approach.

Figure 1 shows the fraction of establishments which reported the number of trained workers and the number of trainings, respectively. Beginning in 1997 the proportion of establishments that refer to the number of trainings was close to 25 percent, decreased over the course of time and settled down at roughly 17 percent. The decrease in the proportion is probably because the mean establishment size in the sample of the IAB Establishment Panel also reduces over the years and larger establishments tend to report the number of trainings more often (Figure 2).

In order to get a first idea of what might happen if one relies only on establishments that report the number of trained workers we run a linear probability model with a dummy variable taking the value one if the establishment reports the number of trained workers and zero if the establishment reports the number of trainings as the dependent variable (see Table 2). As independent variables we include firm size, industry and a dummy for West Germany (column 1) and further establishment specific variables (see column 2). It turns out that larger establishments are clearly more likely to report the number of trainings. What is more, non-profit establishments tend also to report rather the number of trainings while establishments from the production and service sectors are more likely to report the number of trained workers. Finally, establishments located in Eastern Germany are more likely to report training courses.

Furthermore, our second specification reveals that the reporting mode also hinges on the presence of works councils and firms' innovation activity. Thus, it can be concluded that relying only on firms that report the number of trained workers might lead to biased results if one controls not sufficiently for the selection mechanism.

3 Idea by Düll and Bellmann (1998) and limitations

Düll and Bellmann (1998) suggest a simple but effective approach to convert the number of trainings to the number of trained workers at the establishment level. First, one needs to cluster similar establishments. The underlying assumption is that these (relatively) similar establishments are also comparable in their further training activities. Next, one calculates a ratio using the information on the number of trained workers and the number of training courses within each cluster. Finally, the ratio is used as a weighting factor to calculate the number of trained workers for the establishments that provide the number of trainings.

We now describe the procedure in more detail and outline the limitations.

$$\omega_g = \frac{\frac{P_g}{B_{P_g}}}{\frac{T_g}{B_{T_g}}}$$

The weight ω_g is calculated within different groups g. Düll and Bellmann (1998) suggest groups based on industry, establishment size, and East/West location.[3] The numerator comprises the sum of trained workers within a group g (P_g) divided by the sum of all workers of establishments within the same group g that provide the number of trained workers (B_{pg}), i.e. share of trained workers in group g of establishments that provide the number of trained workers. In the same way, the denominator comprises the sum of trainings (T_g) of those establishments that provide the number of trainings relative to the sum of all workers in the respective establishments (B_{Tg}). Thus, ω_g gives the ratio of average training intensities within a group of (relatively) similar establishments. The weight ω_g is used to convert the number of training courses to the number of trained workers at the establishment level by multiplying the former with ω_g.

However, while this method looks straightforward at first glance, some pitfalls have to be taken into account. First, the measure requires a sufficient number of observations of both the number of trained workers and the number of training courses within each group in order to calculate a meaningful weight ω_g. If the groups are too small, the likelihood increases that either no establishments providing the number of trained workers (P_g) or no establishments providing the number of trainings (T_g) are in a group g. If P_g is missing in a group it is impossible

[3] Düll and Bellmann (1998) use 15 industries, 3 establishment size categories (1-99, 100-499, ≥ 500) and East/West. We discuss the choice of categories below.

to calculate ω_g for that group g and thus one cannot convert the number of trainings for establishments in such groups.[4]

A second limitation of the approach is that there could be cases where the converted number of trained workers is higher than the reported number of trainings, which is implausible as the number of reported trainings logically limits the number of trained workers to the maximum of the number of participations. This issue arises for all those establishments with a calculated weight larger than 1.

Another limitation is that the converted number of trained workers could be larger than the establishment size. First, as mentioned above, this problem may arise as the training information refers to a period while establishment size refers to a reference date. Hence, it could be that some establishments report more trained workers than workers employed at the reference date. Second, there are also cases where the converted number of trained workers is implausibly high compared to the respective establishment size. On the one hand, this might happen if the weight ω_g is calculated very imprecisely. On the other hand, this might also be due to misreporting of the training information by the survey respondents. Empirically it is not possible to disentangle both errors and it is not possible to define a critical threshold (unlike to the above mentioned limitations). Hence, one has to rely on plausibility considerations if the converted number of trained workers is much higher compared to the establishment size. We will come back to these limitations in the next section.

4 Implementation

In what follows, we implement and discuss six different aggregation levels and finally propose one classification for grouping the establishments. However, the researcher is free to choose other aggregation levels. This might be even more relevant if one wants to apply this method to subgroups of employees (e.g. number of trained workers with low qualification) with fewer observations.

4.1 Aggregation levels for the weighting factor

The first column of Table 3 provides an overview of the different aggregations analyzed and discussed in this paper. We always use East/West location and the survey year to form the groups (as in Düll and Bellmann 1998). However, we vary the industry and establishment size dimensions, as reported in the very first column of Table 2. We use 3, 5, or 10 establishment size categories and 4 or 12 industry categories (see Appendix Table A1). We aggregate our industry classes from the categories of the sampling scheme of the IAB Establishment Panel. We account for changes over time in the industry classification when grouping our variables.[5]

[4] If T_g is missing, ω_g also cannot be calculated, but within groups that comprise no establishments providing training courses conversion to number of trained workers is not necessary.

[5] The overall results are not sensitive to the choice of how to (qualitatively) aggregate the establishment size and industries but rather regarding the number of aggregates. We therefore vary the number of size and industry categories in our different versions.

The challenge is to define a grouping level in order to calculate the weight ω_g with comparable establishments and at the same time to have a weight with as few as possible limitations. The six different versions are sorted by accuracy: The more categories of establishment size and industry are used, the more similar are the establishments within each group, and the more groups are generated (column 1) accompanied by fewer establishments within each group (column 2). Furthermore, the share of observations within groups that report the number of trainings varies across the different versions (column 3). We expect that establishments in smaller groups are more comparable. Thus our method should be more precise. However, smaller groups come at the cost of other drawbacks, which we explain below.

4.2 Discussion

The first limitation we discussed above is an insufficient number of observations within each group to calculate ω_g. A necessary condition is that there are as few as possible groups with only establishments reporting the number of trainings (i.e., no establishments reporting the number of trained workers, missing P_g.). Column 4 of Table 3 provides the number of groups and observations without establishments reporting the number of trained workers for each version. The smaller the average group size (the more groups are used) the more likely are groups with missing P_g. For example, in version 6 we use 2,056 groups to calculate ω_g (column 1), but overall there are 62 additional groups without P_g (column 4).[6] Here, a broader level of groups reduces this error.

As a second limitation, we pointed out that the approach might lead to groups with $\omega_g > 1$, which is implausible. In other words, $\omega_g = 1$ holds if each trained worker gets only one training, but – in order to be trained – one reported training course is the minimum. Column 5 of Table 3 reports the number of observations if the converted number of trained workers exceeds the number of trainings. Column 5 also includes the share of these observations relative to all converted observations. In our versions the share is always smaller than 10 % of observations but increases if the groups are smaller, i.e. observations within the groups are more similar. Here, a broader level of groups clearly reduces this error.

Finally, we highlighted that it might be suspicious if the number of trained workers exceeds employment at the reference date dramatically.[7] Column 6 of Table 3 reports by how much the converted number of trained workers exceeds the establishment size (deciles 1, 5 and 9 of the distribution) and the respective number of observations. E.g., in version 2 half of all establishments with converted training information report less than 33.4 percent more trained workers in the first half of the year compared to the reported employment at the reference date. Here, a more precise level of groups reduces this error to some degree.

[6] Additionally, version 6 includes 241 groups with no T_g (not reported).

[7] Here, we can compare the establishments reporting the number of trained workers and their establishment size with the converted training information of the other establishments. Although, at the median, the share of trained workers is 95.8 % larger than the establishment size, the number of observation with this error is small (32). Second, this miscalculation could also arise because the weight ω_g is imprecise. However, we do not have a sensible measure to predict the source and size of the error.

5 Guidelines for users

In this section, we recommend one of our versions to aggregate the establishments in (relatively) similar groups to calculate ω_g with as few as possible limitations. In a next step we use the converted number of trained workers to show some descriptive results.

As there are two limitations that indicate more problems when the groups are small we recommend choosing rather a more aggregated version. What is more, the first two limitations offer also clear thresholds while for the third limitation one has to rely solely on plausibility considerations. Thus we suggest to use **version 2** with five establishment size and four industry categories per year and region. The accompanying do-file is based on this aggregation.

Figures 3a/b depict the resulting (average) number and share of trained workers for all establishments after converting the number of trainings and compares these values to establishments that report only the number of trained workers. First, the number of trained workers increases roughly by one third. Most important, starting in 2014 all establishments could report the number of trained workers only. As the mean number of trained workers is roughly the same as in 2013 (after converting) we are quite confident that our procedure delivered meaningful results. Looking at the share of trained workers we finally find that the conversion procedure does not change the overall picture substantially. We take the fact that both shares are quite similar as cautious evidence that our approach works.

We recommend users either to carefully take possible selection mechanisms between establishments that report the number of workers and those that report trainings into account or to rely on a conversion mechanism like ours. Given the method and its limitations we suggest to follow the subsequently discussed guidelines.

First, given the sensitivity regarding group size we recommend the researcher to implement our code for the *whole* IAB Establishment Panel and to use the generation of a panel dataset as proposed by Umkehrer (2017). Of course, it is possible to drop parts of the data that define our groups (e.g. running the code only for one year or only for West Germany) without reducing the quality of the procedure. However, users should avoid running the code only on their (smaller) analysis sample as this constrains the procedure needlessly. Please also note that industry classification changed over time.

Second, we urge users to perform a set of robustness checks to learn about the sensitivity of the results. To start with, it is helpful to compare results based on samples with and without converted training information.

Third, another check should ensure that results are not sensitive with regard to $\omega_g > 1$. To do this, one could simply drop such observations or restrict the value ω_g to 1.

Fourth, it may be worthwhile to test if results depend on establishments that seem to train excessively (e.g. when their current employment is lower than the number of trained workers).

Finally, we invite users to extend the approach to the training information for the different subgroups and to share their code with the research community.

6 How to work with the do-file

Users should consider the following points when working with the do-file `training_conversion_FDZMR.do` :

- Download of the do-file

  Users can download the do-file from the FDZ homepage: http://doku.iab.de/fdz/reporte/2018/MR_02-18_EN_programs.zip.

- Use within FDZ environment

  When working on-site or via remote-execution with JoSuA no directories have to be specified. The program automatically assumes a directory structure as set up in the FDZ-Gästenetz and JoSuA, respectively, and defines the associated macros.

  To execute the program for panel generation the following line

  ```
  do “$prog\training_conversion_FDZMR.do"
  ```

  has to be written in the *master.do*. In case of remote execution both *master.do* and `training_conversion_FDZMR.do` have to be uploaded to JoSuA.

- Working with test data

  The program also runs with the test data of the IAB Establishment Panel.

- Set globals

  By default, data set and variable names are as in Umkehrer (2017). Thus the globals have to be defined by the user if the IAB Establishment Panel is not generated using the generation of a panel dataset as proposed by Umkehrer (2017).

- Generated files

  The program saves a log-file documenting the programing to the directory ‘log’ and a .dta-file including the weight (w), idnum and year to the directory ‘data’. By default, both log-file and data-file are named in exactly the same way as the do-file.

- Merge weight and calculate new training variable (in own analysis data set)

  At the very end of the do-file we show how to merge w and convert the training information, which can be done in the analysis data set.

Tables

Table 1: Availability of training information

period	training indicator	# of trained workers	# of trainings	survey interval	conversion possible
1993-1995	yes	yes	no	annually	does not apply
1997-2006	yes	yes	yes	biennially	yes
2007-2013	yes	yes	yes	annually	yes
Since 2014	yes	yes	no	annually	does not apply

Table 2: OLS Regression on indicator “trained workers (1) vs. training courses (0)”

	Specification 1 (1)	Specification 2 (2)
Establishment size (ref: small)		
50-100	-0.062***	-0.046***
	(0.005)	(0.005)
100-250	-0.130***	-0.108***
	(0.005)	(0.006)
250-500	-0.197***	-0.170***
	(0.007)	(0.008)
> 500	-0.331***	-0.300***
	(0.008)	(0.009)
Industry (ref.: Agriculture, mining, electricity)		
Manufacturing and construction	0.032***	0.030***
	(0.009)	(0.009)
Service, trade, hotels and restaurants	0.022***	0.020**
	(0.009)	(0.009)
Non-profit organization,	-0.059***	-0.049***
public administration	(0.011)	(0.011)
West (d)	0.015***	0.015***
	(0.003)	(0.004)
Age: younger than 5 years (d)		-0.001
		(0.004)
Works council (d)		-0.030***
		(0.004)
Collective bargaining agreement (d)		-0.001
		(0.003)
New technical facilities (d)		0.006**
		(0.003)
Single establishment (d)		0.008**
		(0.004)
Innovation (d)		-0.013***
		(0.003)
Further controls	no	yes
Years	yes	yes
N	102,183	102,183

Notes: Further controls: workforce composition. Standard errors are clustered as robust (cluster: establishment) and are shown in parentheses. Significance levels: $*$ $p < 0.10$, $**$ $p < 0.05$, $***$ $p < 0.01$.

Table 3: Aggregation levels for the weighting factor

Versions		Number of groups	Group size: mean (s.d.) [median]	Within group share of observations with number of trainings (average)	No. of groups (observations) with only information on the number of trainings but no information on the number of trained workers (i.e., groups with missing ω_g because of missing P_g)	Observations with trained workers > number of trainings (share in all converted observations)	No. and distribution of establishments with trained workers > establishment size
		(1)	(2)	(3)	(4)	(5)	(6)
1	3 establishment size categories 4 industry categories	288	392.63 (555.26) [201.5]	0.28	0 (0)	635 (3.3%)	Obs. 1,120 10% -149% 50% -36.6% 90% -5.3%
2	5 establishment size categories 4 industry categories	480	235.58 (369.80) [125]	0.26	0 (0)	699 (3.6%)	Obs. 1,070 10% -141.3% 50% -33.4% 90% -5.6%
3	10 establishment size categories 4 industry categories	865	130.22 (145.86) [79]	0.24	22 (50)	799 (4.2%)	Obs. 966 10% -119.9% 50% -30.4% 90% -4.4%
4	3 establishment size categories 12 industry categories	729	154.92 (178.48) [94]	0.25	5 (10)	1,457 (7.6%)	Obs. 1,190 10% -147.2% 50% -33.7% 90% -3.7%
5	5 establishment size categories 12 industry categories	1208	93.30 (118.28) [47]	0.23	9 (15)	1,608 (8.3%)	Obs. 1,083 10% -137.2% 50% -31.3% 90% -4.9%
6	10 establishment size categories 12 industry categories	2,056	53.466 (52.567) [35]	0.21	62 (126)	1,857 (9.7%)	Obs. 906 10% -101.0% 50% -26.4% 90% -3.9%

Figures

Figure 1: Share of establishments reporting the number of trainings or the number of trained workers

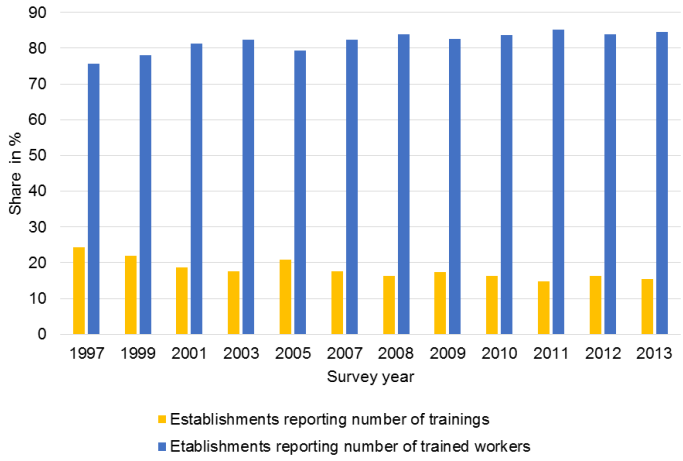


Figure 2: Share of establishments reporting the number of trainings or the number of trained workers by establishment size (1997-2013)

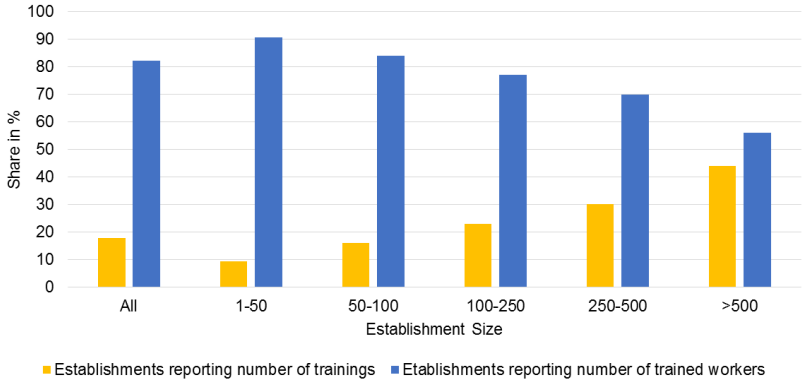
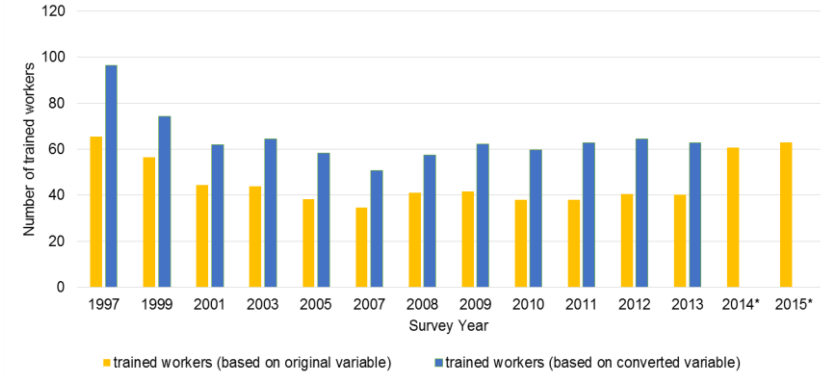
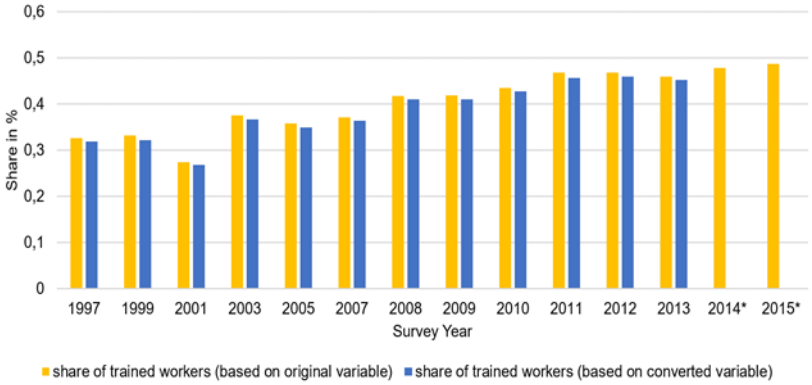


Figure 3: Comparison of converted and original information on the number and share of trained workers

a) Number of trained workers



b) Share of trained workers



* indicates survey years were respondents could report only the number of trained workers

References

Bechmann, S., Tschersich, N., Ellguth, P., Kohaut, S., & Baier, E. (2017). Technical report on the IAB Establishment: panel wave 23 (2015). FDZ-Methodenreport, 08/2017.

Berg, P. B., Hamman, M. K., Piszczek, M. M., & Ruhm, C. J. (2016). The relationship between establishment training and the retention of older workers: Evidence from Germany. International Labour Review, in press.

Bellmann, L., Bossler, M., Gerner, H.-D., & Hübler, O. (2017): Training and minimum wages: first evidence from the introduction of the minimum wage in Germany. IZA journal of Labor Economics,6(8), 22 S.

Bossler, M., Geis, G., & Stegmaier, J. (2017) Comparing survey data with an official administrative population: assessing sample-selectivity in the IAB Establishment Panel. Quality & Quantity, 1-22.

Düll, H., & Bellmann, L. (1998). Betriebliche Weiterbildungsaktivitäten in West- und Ostdeutschland. Mitteilungen aus der Arbeitsmarkt- und Berufsforschung, 31 (2), 205-225.

Ellguth, P., Kohaut, S., & Möller, I. (2014). The IAB Establishment Panel – methodological essentials and data quality. Zeitschrift für ArbeitsmarktForschung, 47 (1), 27-41.

Fischer, G., Janik, F., Müller, D., & Schmucker, A. (2009). The IAB Establishment Panel things users should know. Schmollers Jahrbuch. Zeitschrift für Wirtschafts- und Sozialwissenschaften, 129 (1), 133-148.

Görlitz, K., & Stiebale, J. (2011). The impact of product market competition on employers' training investments. Evidence from German establishment panel data. De Economist, 159(1), 1-23.

Heywood, J. S., Jirjahn, U., & Pfister, A. (2017). Product market competition and employer provided training in Germany. Research Papers in Economics, Universität Trier, No. 7/17.

Hinz, T. (2016). Personnel policy adjustments when apprentice positions are unfilled: Evidence from German establishment data. Diskussionspapiere, Friedrich-Alexander-Universität Erlangen-Nürnberg, Lehrstuhl für VWL, insbes. Arbeitsmarkt-und Regionalpolitik, No. 99.

Stegmaier, J. (2012). Effects of works councils on firm-provided further training in Germany. British Journal of Industrial Relations, 50 (4), 667-689.

Umkehrer, M. (2017). Combining the Waves of the IAB Establishment Panel. A Do-file for the Basic Data Preparation of a Panel Data Set in Stata. FDZ-Methodenreport 12/2017.

Zwick, T. (2006). The impact of training intensity on establishment productivity. Industrial Relations, 45(1), 26-46.

Appendix

Table A1: Establishment size and industries categories

a) Establishment size in categories

	3 categories	5 categories	10 categories
1	≤ 100	< 50	<5
2	100-500	50-99	5-9
3	>500	100-249	10-19
4		250-500	20-49
5		>500	50-99
6			100-199
7			200-499
8			500-999
9			1000-4999
10			>4999

b) Industries in categories

4 industries	12 industries
Agriculture, mining and electricity and other facilities	Agriculture, mining and electricity and other facilities
Manufacturing and construction	Manufacturing
Trade, services and food services	Construction
Non-profit organizations and public administration	Trade
	Transport and storage, insurances
	Accommodation and food services
	Education
	Human health and social work
	Services*
	Non-profit organizations, public administration

* “services” includes more than one category, depending on the respective stratification variables, see do-file for detailed information.

Imprint

FDZ-Methodenreport 02/2018 (EN)

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

Editorial staff
Dana Müller, Dagmar Theune

Technical production
Dagmar Theune

Download
http://doku.iab.de/fdz/reporte/2018/MR_02-18_EN.pdf

Internet
http://fdz.iab.de/

Corresponding author:
Jens Stegmaier
Institute for Employment Research (IAB)
Regensburger Str. 104
D-90478 Nürnberg
Phone: +49-911-179-3042
Email: Jens.Stegmaier@iab.de