IZA DP No. 7971

The IZA Evaluation Dataset Survey: A Scientific Use File

Patrick Arni
Marco Caliendo
Steffen Künn
Klaus F. Zimmermann

February 2014

The IZA Evaluation Dataset Survey:
A Scientific Use File

Patrick Arni
IZA, DEEP (Lausanne) and CAFÉ (Aarhus)

Marco Caliendo
University of Potsdam, IZA, DIW Berlin and IAB

Steffen Künn
IZA

Klaus F. Zimmermann
IZA and Bonn University

Discussion Paper No. 7971
February 2014

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

ABSTRACT

The IZA Evaluation Dataset Survey: A Scientific Use File[*]

This reference paper describes the sampling and contents of the IZA Evaluation Dataset Survey and outlines its vast potential for research in labor economics. The data have been part of a unique IZA project to connect administrative data from the German Federal Employment Agency with innovative survey data to study the out-mobility of individuals to work. This study makes the survey available to the research community as a Scientific Use File by explaining the development, structure, and access to the data. Furthermore, it also summarizes previous findings with the survey data.

Corresponding author:

Steffen Künn
IZA
P.O. Box 7240
53072 Bonn
Germany
E-mail: kuenn@iza.org

1. **Introduction**

In modern welfare states, active labor market policies (ALMP) such as job search assistance, training programs, public employment programs and wage subsidies are intended to reintegrate the unemployed back into the labor market. Given that countries spend significant shares of their budgets on activation measures (see OECD, 2013), it is important for policy makers to ascertain if such programs indeed improve the labor market prospects of participants. In order to obtain reliable estimates for the impact of ALMP and understand why and how programs work or not, both appropriate econometric methods and suitable data are required. While the development of econometric methods and computational power has increased dramatically during recent decades, data availability or the information content of existing datasets still represent a bottleneck.

To overcome the problem of data limitations within the field of labor economics, IZA has recently implemented a large-scale survey, the *IZA Evaluation Dataset Survey* (IZA ED Survey). In contrast to population-representative surveys, this survey has the advantage that it captures a large entry sample of unemployed individuals and therefore includes large shares of participants in ALMP programs. In fact, the IZA ED Survey covers a panel of 17,396 individuals who registered as unemployed at the Federal Employment Agency in Germany between June 2007 and May 2008.[1] Based on computer assisted telephone interviews (CATI), the individuals were interviewed up to four times. Starting at their entry into unemployment, the individuals were interviewed at frequent

intervals during the first 12 months of unemployment and in the long-run after three years.

This data allows the researcher to observe dynamics with respect to individual and labor market characteristics during the early stage of unemployment, as well as tracking long-run outcomes. Within the survey, information on labor market activities, ALMP participation, migration background, search behavior, ethnic and social networks, psychological factors, cognitive and non-cognitive abilities, attitudes and preferences was recorded. Its large sample size of individuals entering unemployment, in combination with its broad set of variables and the measurement of unemployment dynamics (due to several interviews during the first three years after unemployment entry), offers new perspectives for empirical labor market research. Besides the evaluation of ALMP programs, this dataset provides a good empirical base to investigate all aspects of the transition process from unemployment to employment. In particular, the combination of rich information on individual characteristics and longitudinal data allows designing detailed studies concerning the interplay of personal (search) behavior and attitudes, labor market outcomes and labor market policies.

The IZA ED Survey is now available as a Scientific Use File. This paper introduces the concept of the Scientific Use File to the scientific community by illustrating the background and motivation for the creation of this dataset in Section 2, before explaining the development, structure and access to the data in Section 3. In Section 4, we provide an overview of applied studies that have used

this dataset in the past, and provide some ideas on further possible fields of application and an outlook in Section 0.

2. **Background**

The starting point for the creation of the IZA ED Survey is based on the aforementioned existence of data limitations in the field of program evaluation. As a first step to overcome such limitations and obtain empirical evidence on the effectiveness of labor market policies, many European countries have recently opened their administrative databases for scientific research. The advantages of administrative data are straightforward: they are consistently and accurately collected, resulting in highly reliable data covering a large number of observations (in some cases even 100% of the population). They are regularly updated such that long time periods are observable usually and the specific use of ALMP programs is directly visible. In addition, the provision of administrative data for scientific research reflects a cost-effective way of providing highly reliable and representative data, as these data are collected for administrative purposes anyway.

However, there are also some limitations associated with administrative data, reducing its usefulness for scientific purposes. Besides a very restrictive access due to data security issues, given that administrative data are collected for administrative purposes the range and variety of variables is quite restricted. Important variables for scientific research such as social networks, personality

traits, cognitive skills, attitudes or ethnic identity are usually not important for administrators and hence are not included in administrative databases.

However, recent studies have shown the high relevance of such variables in empirical studies in the field of labor economics (e.g. Borghans et al., 2008, Bonin et al., 2007, Constant and Zimmermann, 2008). Further information that is also needed for labor market research yet not included in administrative data includes, for instance, information on job search behavior, such as reservation wages, search intensity or search channels, or job satisfaction and individuals' expectations concerning their future labor market success and health condition. Indeed, such information is crucial towards understanding *why* certain ALMP programs work and others do not. Thus, survey data are needed to answer fundamental research questions that cannot be answered by using administrative data.

In order to provide a base for empirical research on such questions of social behavior, many countries have started initiatives to create survey data for scientific purposes. The most known surveys are generally the large population-representative surveys such as the *German Socio-Economic Panel* (GSOEP), the *Current Population Survey* (CPS) in the U.S., the *British Household Panel Survey* (BHPS) or the recently started *Household, Income and Labour Dynamics in Australia Survey* (HILDA). Such surveys are widely used and depict the main workhorse in empirical social sciences.

However, they cannot solve the data restrictions within specific research areas such as the evaluation of ALMP programs, the economics of migration or education. In these areas, population representative surveys are not particularly appropriate as they capture insufficient information and sample sizes concerning certain subgroups of the population (e.g. job seekers, immigrants, pupils) or with respect to specific subjects (e.g. unemployment, migration aspects, school performance).

To overcome such data limitations, several institutions have started data initiatives to abolish particular data restrictions within certain research areas. For instance, the *New Immigrant Survey* in the US has been implemented to create a data base for analyzing policy questions on immigrants in the U.S. (see Guillermina et al., 2000). Consistently, the *Rural-to-Urban Migration Dataset* was created to analyze the massive migration flows from rural to urban areas in China (see Kong, 2010; Akgüc et al., 2013). Moreover, topic-specific surveys have also been implemented, e.g. the *German Panel Analysis of Intimate Relationships and Family Dynamics* (see Huinink, 2011) to investigate mechanisms of intergenerational transmission or the *German National Educational Panel Study* (NEPS, see Blossfeld et al., 2011) to analyze questions within the field of economics of education.

In line with this strand of data projects, IZA has recently implemented the IZA ED Survey on unemployed individuals. The main aim of this survey is to generate an optimal data base for the evaluation of social and labor policies, as

well as studying the transition process from unemployment back to employment. Therefore, the underlying population of the survey focuses solely on entries into unemployment, given that such individuals are primarily targeted by labor market policies. The survey is now available as a Scientific Use File, which will be distributed by the *International Data Service Center* (IDSC) of IZA.[2]

A distinctive and attractive feature of the IZA ED Survey is that it can be merged to administrative data as provided by the Institute for Employment Research (IAB) in Nuremberg, the research institute of the Federal Employment Agency (see Caliendo et al., 2011a for details). The administrative data cover daily information on individuals' labor market activities, including wages and benefits, for a period covering from 1975 until present. The merging of the IZA ED Survey with the administrative data provides the additional advantage of combining the variety of survey information with the high reliability and large observation window of the administrative data. However, the administrative data are subject to very restrictive data security legislation that currently prevents public access to the merged dataset. IZA is actively engaging in joint work with the IAB to find a solution that will provide access to the merged dataset in the future.

3. **The Data**

The aim of the IZA ED Survey was to interview new entries into unemployment, collecting detailed information on these individuals and their

labor market activities, starting at entry into unemployment until three years after. The following section describes the underlying target population, the construction of the survey, the questionnaire and characteristics of the finally realized samples, as well as providing guidance on data access. Thereby, the focus is solely on the main features of the data. A very detailed and more technical description of the data construction, including a description of the questionnaire, an extensive analysis of non-response and panel attrition, and the calculation of panel weights can be found in the User Manual of the IZA ED Survey.[3]

3.1. The Target Population and Sampling

The IZA ED Survey consists of individuals who registered as unemployed at the German Federal Employment Agency within the period from June 2007 to May 2008.[4] The aim was to construct a sample of "new" entries into unemployment, i.e. prime-age individuals who enter unemployment, are looking for a job and are eligible to participate in ALMP programs.

TABLE 1: APPLIED SAMPLE RESTRICTIONS

Pre-interview restrictions applied to the sample drawn from the unemployment inflow statistic
1) Age restriction: 16-54 years at entry into unemployment
2) Exclusion of unemployment benefit type II recipients
3) Exclusion of re-entries into unemployment after a period of sickness or participation in ALMP programs
Restrictions during the interview
4) Verification of unemployment entry and previous activities by respondents
5) Exclusion of "pseudo" unemployment entries: Individuals who signed a contract for a new job already at entry into unemployment and hence do not search for employment

The contact information on individuals entering unemployment was drawn from the monthly unemployment inflow statistic of the Federal Employment Agency. This statistic records individuals when they register as unemployed at the Federal Employment Agency – if eligible to unemployment benefit type I – or the agency responsible for the unemployment benefit type II. While unemployment benefit type I is paid to individuals who made contributions to the unemployment insurance in the past, unemployment benefit type II is a means-tested, tax-funded benefit that is paid to long-term unemployed or individuals without any previous employment experience (see Konle-Seidl et al., 2010 for an overview on the German unemployment insurance system). Therefore, the unemployment inflow statistic contains a very heterogeneous pool of entries into unemployment, so that – based on the available information included in the unemployment inflow statistic – some restrictions were implemented in order to pre-select the target population (see Table 1 for an overview).

First of all, an age restriction was applied (16-54 years at entry into unemployment) to avoid any influence due to retirement decisions, e.g. individuals might voluntarily enter unemployment in order to retire earlier and bridge the time until the official retirement age. However, given that these individuals are not looking for a job they do not belong to our target population. Moreover, we excluded individuals who received unemployment benefit type II (subject to Social Code II, SGB II) at entry into unemployment, due to three reasons. First, unemployed individuals whose unemployment benefit type I

entitlement elapses after being unemployed for a certain period (in most cases after 12 months) will be technically registered in the unemployment inflow statistic as an entry into unemployment benefit type II. In economic terms, however, this does not represent a new entry into unemployment and thus such individuals should be excluded from the sample. Second, the SGB II records are likely to be incomplete and third, individuals receiving unemployment benefit type II are not eligible to every ALMP program. Therefore, excluding unemployment benefit type II recipients narrows the sample towards the specified target population. As a last step, individuals who are likely to be re-entries into unemployment were excluded. The unemployment inflow statistic technically defines every individual who registers as unemployed after a certain period of not being unemployed as an entry into unemployment. Therefore, periods of sickness or participation in ALMP programs interrupt unemployment spells, so that individuals who did not find a job during that time are counted (again) as entries into unemployment. However, given that these interruptions do not terminate unemployment in economic terms, these spells are not "new" entries into unemployment and thus have to be excluded. Therefore, all individuals who registered as unemployed after a period of sickness or ALMP participation or had an entry into unemployment in the previous month were excluded.

In addition to the pre-interview sample restrictions, a very detailed screening took place at the beginning of each interview in order to finally identify the target population. This verification procedure was required as the available information

provided by the unemployment inflow statistic only allowed for a raw identification of the target population. First of all, each individual had to answer several questions about his/her current unemployment entry to ensure that the individual unambiguously belongs to the pre-defined target population. Most importantly, as this is not observed in the unemployment inflow statistic, individuals who reported having already signed a contract for a new job at entry into unemployment were dropped, as they are not searching for employment.

This two-step procedure combining the pre-interview sample restrictions and the screening during the interview guarantees that only individuals who unambiguously belong to the specified target population were interviewed.

3.2.   Construction of the Survey and Response Rates

The IZA ED Survey is constructed as a panel where individuals entering unemployment within the period from June 2007 until May 2008 were interviewed at least three times, i.e. at entry into unemployment, as well as 12 and 36 months later (see Figure 1). In addition, three selected monthly cohorts, i.e. entries into unemployment in June and October 2007, and February 2008, received an additional interview six months after entry into unemployment. The main aim of this interim wave is to measure dynamics with respect to changes in individual and labor market characteristics during the early stage of unemployment. Due to restricted financial means and the risk of higher panel attrition for these individuals, the interim wave was restricted to three cohorts only, distributed over the entire year to avoid any bias due to seasonality.

FIGURE 1: STRUCTURE OF THE SURVEY
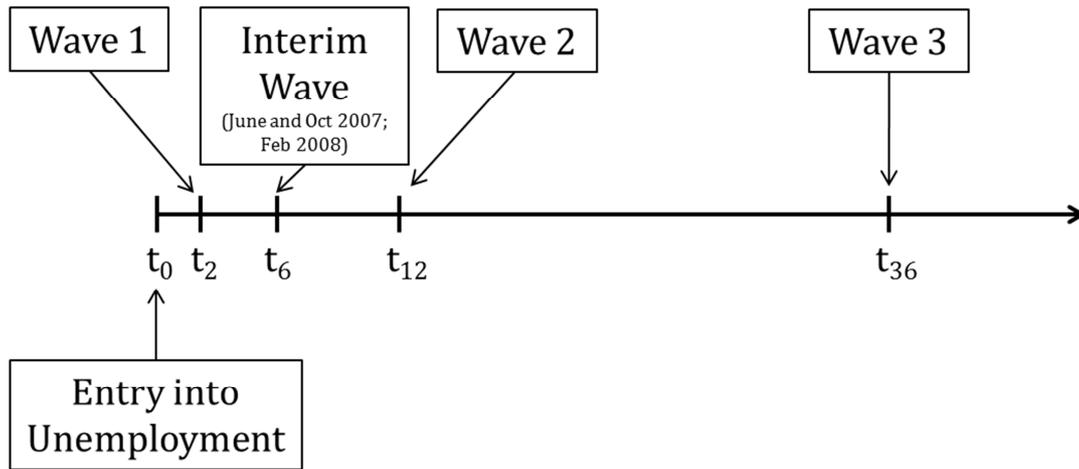
The interviews were performed by means of pre-tested computer assisted telephone interviews (CATI), conducted by a professional survey institute[5]. In advance of the interview, each individual received a letter prior to being contacted. The main aim of the letter was to increase the acceptance of the study and therefore participation rates by informing individuals about the content and background of the survey, as well as data security legislation. The interviews were held in German and, for the two most important immigrant groups in Germany – Russians and Turks – in their native language, if German language skills were insufficient.

As explained above, the contact information for potential interview respondents was provided by the unemployment inflow statistic of the German Federal Employment Agency, which records individuals entering unemployment on a monthly basis. Within the period of interest (May 2007 to June 2008), the inflow statistic recorded around eight million entries into unemployment. In order to interview each individual as immediately as possible after entry into

unemployment, the survey was implemented on a monthly basis. At the end of each month, a random sample of new entries into unemployment was drawn from the unemployment inflow statistic (following the sample restrictions as depicted in Table 1) and immediately delivered to the survey institute.

Subsequently, the survey institute prepared the data for the interview and contacted the individuals in order to conduct an interview. In total, 81,399 addresses were available for the first interview. The data generating procedure, i.e. sample preparation, transfer to the survey institute and contacting of individuals, was successfully implemented within an average of only two months, so that the respondents received the first interview closely after entry into unemployment (indicated by t_2 in Figure 1). In subsequent interview waves, only individuals who agreed during the first interview to participate in subsequent waves were contacted again. Individuals who dropped out once were not contacted again, i.e. only respondents in wave 2 were contacted for an interview in wave 3.

Table 2 provides an overview of the finally realized interviews in each wave and sample. The upper part shows the numbers for the full sample, while the lower part provides a separate overview for the restricted sample only (three selected monthly entry cohorts). The objective for the first interview wave was to realize around 1,500 interviews each month, totaling approximately 18,000 interviews. It can be seen in the upper part of Table 2 that this goal was almost accomplished with 17,396 interviews realized in the first interview wave,

whereby 90.8% agreed to participate in the panel. Based on these 15,802 observations, 8,915 interviews could be finally conducted in the second and 5,786 in the third wave, which corresponds to 51.2% and 33.3% of the initial sample. For the restricted sample, i.e. the three selected entry cohorts who also had an interim interview six months after entry into unemployment, 4,423 interviewees were available in the first interview wave, 2,548 in the interim, and 1,589 and 985 in the second and third wave, respectively. Panel attrition here is slightly higher than in the full sample, which is most likely due to the additional interview.

TABLE 2: NUMBER OF OBSERVATIONS

Full sample				
	Wave 1		Wave 2	Wave 3
	Realized	Willing to participate in the panel	Realized	Realized
Number of interviews	17,396	15,802	8,915	5,786
%	100	90.8	51.2	33.3
%	---	100	56.4	36.6

Restricted sample: Three selected entry cohorts (June and October 2007, February 2008)					
	Wave 1		Interim Wave	Wave 2	Wave 3
	Realized	Willing to participate in the panel	Realized	Realized	Realized
Number of interviews	4,423	4,060	2,548	1,589	985
%	100	91.8	57.6	35.9	22.3
%	---	100	62.8	39.1	24.3

3.3. Non-Response and Panel Attrition

Collecting data by a telephone survey bears the risk that the implementation of the survey introduces a selection bias, as individuals are free to choose whether or not to participate. Such a selection bias might arise due to selective non-response behavior at the first interview and attrition in later interview waves. An initial non-response bias occurs if the first interview can only be realized for a selective subsample of the underlying population, which will introduce a selection bias if the non-response is correlated with individual characteristics. Panel attrition occurs if individuals are willing to give an interview in the initial wave but drop out and do not return in subsequent interview waves, e.g. due to subsequent refusal, death, relocation or associated problems for tracing individuals. Similar to non-response, panel attrition will introduce a selectivity bias in the sampling if drop-outs are systematically correlated with individual characteristics. If one can credibly assume that selectivity is mostly driven by characteristics that are observed, the potential selection bias can be rebalanced by a weighting scheme.

In order to reveal whether the implementation of the first interview finally led to a representative sample of the target population, it would be necessary to compare characteristics of individuals who participated in the first interview wave with those of the underlying target population. Another possibility is to compare individuals who were contacted but refused to give an interview with

survey participants. Both comparisons would answer the question of whether the realized sample suffers a non-response bias.

However, in the case of the IZA ED Survey, the final identification of the target population took place during the interview. This was necessary as some important screening characteristics are not observable in the unemployment inflow statistic, and thus individuals had to be contacted in order to finally verify whether or not they belong to the target population. As a consequence, the sample extracted from the unemployment inflow statistic and the sample of interview refusals still contain individuals who are not part of the target population. This actually prevents us from running a representative non-response analysis for the first interview wave. For instance, if we detected differences between interview refusals and survey participants, we could not conclude that such differences are driven by selective non-response behavior given that the group of refusals still contains individuals who are actually not eligible for an interview.

This is a common problem with telephone surveys where the final identification of the target population takes place during the interview. What is usually undertaken in such cases is to provide as much information as possible concerning the data generation process. We therefore provide a descriptive comparison of survey participants with the sample extracted from the unemployment inflow statistic and interview refusals with respect to observable characteristics in Table 3.

TABLE 3: COMPARISON OF GROSS SAMPLE, REFUSALS AND REALIZED SAMPLE IN WAVE 1

	Gross Sample	Wave 1		p-value	
		Refusals	Realized Sample		
	(1)	(2)	(3)	(1) vs. (3)	(2) vs. (3)
Number of observations	81,391	5,388	17,396		
Female	43.9	44.5	47.4	0.000	0.000
Age category					
≤ 24 years	28.0	28.0	27.6	0.263	0.581
25 to 34 years	26.6	25.7	26.1	0.114	0.631
35 to 44 years	24.6	25.7	25.1	0.308	0.281
≥ 45 years	20.7	20.5	21.3	0.061	0.217
German citizen	91.1	92.6	92.7	0.000	0.001
School degree					
None, unknown	8.0	6.6	5.4	0.000	0.001
Lower secondary school	34.7	35.2	30.5	0.000	0.000
Middle secondary school	36.5	37.3	37.6	0.007	0.666
Advanced middle sec. school	7.6	8.0	9.3	0.000	0.003
Upper secondary school (A-level)	13.2	12.9	17.1	0.000	0.000

Note: Numbers are percentages and based on administrative information included in the unemployment inflow statistic. Gross sample: Sample extracted from the unemployment inflow statistic (excluding eight individuals due to missing information in observable characteristics). Refusals: Individuals who have been contacted and refused to give an interview but were willing to provide at least some information about their current labor market activities (so-called soft-refusals). P-values are based on a simple t-test of equal means.

It can be seen that the realized sample in wave 1 differs from the two other samples in terms of observable characteristics. We find that women, natives and individuals with higher school attainment have a higher probability of participating in the survey. Although the differences are small, they are mostly statistically significant (as indicated by respective p-values). However, as explained above, we do not know whether these differences arise due to selective non-response behavior or because the gross sample and the refusals still contain individuals who do not belong to the target population. Therefore, we decided to follow different experts in the field of survey design and refrain from providing weights to correct for these differences.[6]

Assuming that the realized sample in the first interview wave is a random sample of the underlying target population, in a second step we assess whether attrition in subsequent interview waves introduces a selection bias. Given that only a small subgroup of the initial sample remains in the survey until the third interview (around 33%, see Table 2), it is likely that panel attrition is correlated with certain individual characteristics. Therefore, we compare individuals in the first wave to those who also participate in later waves. We find that women, natives, better educated and older individuals, as well as those with more employment experience and higher earnings in the past are more likely to remain in the survey. Intuitively, we also find that individuals who faced communication problems during the first interview are less likely to give an interview again. Therefore, the analysis of survey drop-outs confirms that panel attrition in the IZA ED Survey is systematically correlated with observable characteristics. Panel weights are provided with the data in order to correct for selective panel attrition (see user manual for details).

3.4. The Questionnaire

Table 4 provides an overview of the general structure of the questionnaire and a list of variables included in each wave. It can be seen that the majority of questions are included in each wave, so that the information was updated at different points in time (see Figure 1). Note that the list of variables only depicts a crude summary of the rich content of the survey, with each category indicated

in Table 4 represented by several questions in the questionnaire (see Section 3.6

for access to the questionnaires).

TABLE 4: CONTENT OF THE SURVEY

Variables	Wave 1	Interim Wave	Wave 2	Wave 3
Cross-sectional information				
Information on the initial unemployment entry	x			
Individual characteristics (e.g. age, sex, region etc)	x	x	x	x
Migration and social background	x	x	x	x
Language skills	x		x	x
Education	x			
Personality (Big-5, Locus of control)	x	x	x	x
Intergenerational transmission	x	x	x	x
Labor market networks	x	x	x	x
Household composition	x	x	x	x
Household income	x	x	x	x
Debts	x		x	x
Life satisfaction	x	x	x	x
Job search and reservation wage	x	x	x	x
Role of Employment Agency (job search)	x	x	x	x
Details on placement/education voucher	x	x	x	x
Benefit receipt and sanctions	x	x	x	x
Labor market activity at interview		x		
Participation in ALMP		x		
Interview-specific information (e.g. date, language)	x	x	x	x
Willingness to compromise during job search[a]	x	x	x	x
Motivation to contact Employment Agency[a]	x			
Health and physical condition[a]	x	x	x	x
Emotional and psychological conditions[a]	x	x		
Drinking and smoking behavior[a]	x	x	x	x
Change of labor market networks during unemployment[a]	x	x	x	x
Personality (risk, trust, patience, reciprocity)[a]	x	x	x	x
Cognitive tests[a]	x	x	x	x
Daily activities and routines[a]	x	x	x	x
Personal appearance[a]	x	x	x	x
Longitudinal information on labor market activities				
Dependent employment	x		x	x
Self-employment	x		x	x
Unemployment	x		x	x
Participation in ALMP	x		x	x
School attendance	x		x	x
Professional training	x		x	x
Internship	x		x	x
Other activities	x		x	X

a) Filled for individuals belonging to the three selected monthly cohorts who also received the interim wave (entries in unemployment in June and October 2007, and February 2008).

The questionnaire consists of cross-sectional and longitudinal questions. The information collected in the cross-section relates to the time of the interview, e.g. 12 months after entry into unemployment in the case of the second wave. Here, individual and job search characteristics are recorded at each interview, which allows the data users to analyze changes over time. As we can see in Table 4, the cross-sectional part records information on the process of entering unemployment, socio-demographics, migration and social background, personality, labor market networks, household and job search characteristics, participation in ALMP programs, the role of the employment agency for job search, life satisfaction and transfer payments.

While such information was collected for all individuals, some questions were only asked to individuals belonging to the three selected entry cohorts that also received the interim wave (entries into unemployment in June and October 2007, and February 2008) in order to measure dynamics in these characteristics during the early stage of unemployment. Here, information is collected concerning an individual's motives to contact the employment agency, his/her willingness to compromise in order to find a job, health, psychical and psychological conditions, drinking and smoking behavior, cognitive skills and additional questions on labor market networks, personality, daily activities and routines as well as personal appearance.

In addition to the cross-sectional questions, the longitudinal section collects monthly information on labor market activities. Therefore, the respondents were asked at each interview (except the interim wave) to update their labor market biography retrospectively, starting at the last interview or, in the case of the first interview, at unemployment entry. Besides recording the labor market activity and its duration in terms of calendar months, very detailed associated information such as earnings, working time or search strategies were also recorded. Ultimately, the longitudinal part allows the data user to reconstruct the complete labor market biography (including spell-specific information) starting at entry into unemployment (t_0) and ending at the last interview in which the individual has participated.

TABLE 5: INTERVIEW DURATION

	Wave 1	Interim Wave	Wave 2	Wave 3
Number of observations	17,396	2,548	8,915	5,786
Average duration of interviews (in minutes)	58	27	41	36

The large amount of information collected by the survey is reflected by the average duration of the interviews, as shown in Table 5, with the first interview taking an average of 58 minutes.[7] The average duration declined in subsequent interviews, which is mainly due to learning effects, i.e. individuals had to answer the same questions several times, as well as a reduction of questions included in subsequent waves (see Table 4). In particular, the exclusion of longitudinal

questions about an individual's labor market activities significantly reduced the average duration in the interim wave.

3.5. **Descriptive Statistics**

Table 6 describes the survey participants, based on information reported in the first interview. It can be seen that 47% of participants are female, 30% are located in East Germany, 40% are married and the clear majority (94%) are German citizens, although 13% are born abroad.

TABLE 6: DESCRIPTION OF PARTICIPANTS IN THE SURVEY

		Survey Participants
Number of observations		17,396
Female		47.4
Age (in years)		33.8
East Germany		29.5
Married		39.8
German citizen		94.2
Not born in Germany		12.5
Upper secondary school (A-level)		20.6
Labor market experience before entry into unemployment		
Share of working lifetime spent in employment		62.9
Last earnings from employment (in €/month, net),	mean 25th centile median 75th centile	1173.9 770 1100 1400
No employment experience		16.0

Note: Numbers are percentages (unless otherwise indicated) and based on the first interview wave.

With respect to labor market activities prior to entry into unemployment, it can be seen that participants spent on average 63% of their lifetime during working age in employment. Among the individuals who were employed at least once in

their working life the median net earnings from their last employment amounted to 1100 Euro/month. Only a minority of 16% had no employment experience at all before entering unemployment.

TABLE 7: DISTRIBUTION OF SELECTED OUTCOME AND TREATMENT VARIABLES OVER TIME

	Total	Wave 1	Interim Wave	Wave 2	Wave 3
Number of observations	17,396	17,396	2,548	8,915	5,786
Labor market status					
Employed (self- or dependent employed)	--	25.1	55.7	62.9	73.4
	--		(55.8)	(60.1)	(72.4)
Unemployed	--	66.6	29.6	23.4	12.9
	--		(29.1)	(24.8)	(13.2)
Education	--	3.3	9.1	7.3	7.1
	--		(9.3)	(8.3)	(7.8)
Others	--	5.0	5.6	6.3	6.6
	--		(5.7)	(6.8)	(6.7)
Affected by labor market policies between interviews[a]					
Participation in active labor market programs	26.3	10.3	33.2	27.9	14.7
			(33.2)	(26.1)	(15.1)
Short-term training	9.4	4.6	16.7	5.5	2.4
			(16.6)	(5.5)	(2.5)
Retraining	8.7	3.3	7.6	8.9	6.3
			(7.4)	(8.3)	(6.5)
Public employment scheme	1.6	0.4	3.6	1.1	1.1
			(4.0)	(1.2)	(1.2)
Wage subsidy[b]	5.0	--	5.5	6.1	4.4
			(5.7)	(5.6)	(4.4)
Start-up subsidy	5.6	2.2	5.5	7.5	1.8
			(5.0)	(6.3)	(1.7)
Received education voucher	8.4	4.6	7.2	9.9	--
			(6.8)	(9.3)	--
Received placement voucher	11.2	5.4	13.1	11.8	--
			(13.0)	(11.6)	--
Sanction in unemployment benefits	8.6	5.0	4.2	7.5	2.5
			(4.5)	(8.6)	(2.7)

Note: Table shows observed values as percentages; weighted values for panel attrition are in parentheses.
a) Share of individuals affected by different policies between current and previous interview (or entry into unemployment in case of wave 1). Numbers for wave 2 refer to the entire period between the first and second interview. Several policies can apply to an individual within the respective time span.
b) Information is not available for wave 1.

In addition, Table 7 shows the distribution of selected outcome variables at each interview. As the implementation of the survey introduced a selection bias due to non-random panel attrition, we provide both the observed and weighted values for subsequent interview waves, calculated using the panel weights that are provided with the data.

First of all, it can be seen that the majority of individuals are able to find employment within the observation window. 25.1% are employed two months after entry into unemployment (at wave 1), increasing to 73.4% after 36 months (at wave 3). Furthermore, it can be seen that the share in unemployment decreases over time, while the share in education is quite stable at around 7-9% (after an initial adjustment).

More interestingly, Table 7 shows the share of individuals who are affected by different labor market policies over time, thus illustrating the high potential of the dataset to evaluate such policies. It can be seen that significant shares of individuals participate in active labor market policy programs, including vocational training, job creation schemes, wage and start-up subsidies, etc. While 10.3% participated in such a program between entry into unemployment and first interview, this increased to 27.9% between the first and second interview. In total, 26.3% of all individuals in the survey participated at least once within the observation window.

The data allow a detailed view on ALMP participation by type of programs. Among the surveyed job seekers, 9.4% participated at least once within the

observation window in a short-term training. This type of programs consists of activities like application training, language courses etc. over a short period of time. The participation rate in retraining – longer-run programs of (re)education – amounts to 8.7%, the one in public employment schemes to 1.6%. The latter program type features publicly sponsored work activities which are not valued by the labor market ("One-Euro-Jobs") and job creation schemes. Wage subsidies and start-up subsidies (to launch self-employment) are assigned to 5% and 5.6% of the individuals, respectively. These participation rates are well comparable to the corresponding figures of the official labor market statistics for the years of 2007 and 2008[8]. Moreover, these rates and the related numbers of observations demonstrate that the IZA ED Survey allows specific treatment effect analyses for different types of ALMP programs separately.

In addition, Table 7 also shows separate numbers with respect to the receipt of education and placement vouchers. These innovative measures have been introduced in Germany in 2003 and are supposed to improve the allocation of training programs (education voucher) and outsource job search assistance to private placement agencies (placement voucher). While previous evaluation studies on education vouchers focused on the effects of voucher redemption (see Rinne, Uhlendorff, Zhao, 2012) due to data restrictions, the IZA ED Survey provides information on both voucher receipt and redemption. This allows a deeper analysis of the education vouchers' effectiveness as an innovative allocation mechanism of ALMP (for example, potential intention-to-treat effects triggered by voucher receipt). Table 7 shows that 4.6% received such a voucher

until the first interview, with this share increasing to 9.9% between wave 1 and wave 2. In total, 8.4% received an education voucher within our sample and observation window.

The survey data also include very detailed information on the receipt of a placement voucher and the resulting job search success, which provides many research opportunities. Here, we observe that 11.2% of the respondents received a placement voucher within our observation window, with 5.4% already receiving a voucher very early during their unemployment spell (reported in wave 1). Later on, the numbers increase to 11.8%, as reported in wave 2.

Besides the participation in a particular program, another key policy that significantly influences the job search behavior of unemployed individuals – in the case that they do not comply with the instructions by the caseworker – is to reduce their unemployment benefits. The IZA ED Survey also includes detailed information on this issue, with Table 7 showing that 8.6% of the individuals were sanctioned at least once within the survey period. Besides the amount and exact timing (announcement, duration) of the sanction, the reason and its subjective assessment by the job seeker are also recorded.

Thus, in sum, the comparative advantage of the IZA ED Survey data is particularly given by the fact that it *combines* rich information about an individual's behavior, attitudes and characteristics with precise and detailed information on ALMP and labor market activities and outcomes. This opens new perspectives for exploring the interactions of these variables.

3.6. **Data Access**

The data are available as Scientific Use Files provided by the IDSC of IZA. In order to acquire more information about how to access to the Scientific Use Files, visit http://idsc.iza.org/iza-ed-survey.

4. **Previous Research Using the IZA ED Survey**

The richness of the dataset provides the basis for a broad set of potential research questions. This can be illustrated using the existing studies with the IZA ED Survey. Table 8 provides an overview of these contributions.

The first strand of studies focuses on the existence of ex ante effects of ALMP programs. Usually, evaluation studies investigate ex post effects on the labor market performance of actual participants. However, the pure announcement of participation in a program might already have an impact on the job search behavior of job seekers. Based on administrative data alone, it is difficult to determine the behavioral mechanics of how ex ante effects operate, given that information on an individual's job search is not included. In contrast, the IZA ED Survey includes information on both the subjective probability of participating in an ALMP program and very detailed information concerning the job search behavior of individuals, such as reservation wages and search channels.

Using this data, van den Berg et al. (2009) find results suggesting that a high perceived participation probability leads to lower reservation wages and

increased search effort. It seems that job seekers try to avoid program participation. The pure announcement of program participation has a "positive" effect on the current job search behavior.

TABLE 8: OVERVIEW OF PREVIOUS STUDIES USING THE IZA ED SURVEY

Nr.	Study	Field/Research Question	Major Finding
1	van den Berg et al. (2009)	Ex ante effects of ALMP participation	Prospect of participating in ALMP programs reduces ex ante reservation wages and increases search effort
2	van den Berg et al. (2011)	Ex ante effects of ALMP participation: Effect heterogeneity with respect to country of origin of migrants	Effects differ considerably by migrant group, probably due to cultural differences
3	Caliendo et al. (2011b)	Role of social networks for job search choices of unemployed job seekers	Individuals with larger networks shift towards more intense use of informal networks and have higher reservation wages
4	Caliendo, Uhlendorff (2011)	Impact of personality and subjective expectations on job search behavior of unemployed individuals	Heterogeneous impacts on job search behavior and transition probabilities to employment
5	Caliendo, Lee (2013)	Impact of obesity on job search behavior and job finding probabilities	Significant impact only for obese women: Lower employment probability and lower wages
6	Krause (2013)	Impact of happiness on job search, job finding probabilities and re-entry wages	Inverse u-shaped relationship between happiness of job seekers and re-employment probability and wages. Happier job seekers exert less search effort.
7	Constant et al. (2011a)	Investigates to what extent the native-migrant gap in economic outcomes can be explained by differences in ethnic identity of migrants and its impact on job search behavior and transition to employment	Less integrated migrants slowly reintegrate into employment, most likely attributable to lower search effort and relatively high reservation wages within this group.
8	Constant et al. (2010)	Analysis of reservation wages of first and second generation migrants	Second generation migrants have higher reservation wages than first generation migrants as they tend to refer to the wage level within the host county, instead of the country of origin
9	Constant et al. (2011b)	Comparison of second generation migrants and natives with respect to the economic impact of attitudes and risk preferences	Differences in attitudes and risk preferences explain lower employment probabilities among second generation migrants

Given that the IZA ED Survey also contains detailed information on migration background, van den Berg et al. (2011) go one step further and run this analysis for different groups of migrants. They find that the ex ante effects differ considerably across migrant groups, most likely due to cultural differences across these groups.

The second strand of studies using the IZA ED Survey concerns the analysis of job search behavior of unemployed job seekers. Besides the evaluation of ALMP programs, this dataset also provides a good empirical base to investigate the job search behavior of job seekers due to the inclusion of several questions about the job search activities of unemployed individuals, such as reservation wages, search channels, willingness to take difficulties to find employment, regional mobility, role of employment agency, etc. The variety of variables included in the IZA ED Survey facilitates studies delivering essential new insights in the field of economics of information and job search.

For instance, Caliendo et al. (2011b) investigate the role of social networks on job search behavior, finding that individuals with larger social networks more commonly use informal search channels and also tend to have higher reservation wages. Moreover, Caliendo and Uhlendorff (2011) discuss how personality traits and (similar to the studies on ex ante effects of ALMP programs) the perceived probability to participate in an ALMP program affect job search behavior and consequently the transition to employment.

Caliendo and Lee (2013) use information on the weight of job seekers to test the hypothesis that overweight individuals behave or are treated differently during job search compared to normal weight individuals. Interestingly, they only find negative labor market effects for overweight women, i.e. lower employment probabilities and lower wages compared to normal weight women. For men, obesity apparently does not alter job search behavior and harm job finding probabilities.

Krause (2013) investigates the influence of individuals' happiness on reemployment probabilities and reentry wage levels of unemployed job seekers. By accounting for the individual's labor market history and information about future job prospects, it was possible to reduce reverse causality bias. The author finds an inverse u-shaped relationship, which means that the optimal level of happiness is not necessarily the highest to maximize reemployment probabilities and wages. The effect on reemployment is driven by the concept of locus of control and the personality traits of neuroticism and extraversion. Interestingly, job search behavior, as measured by the number of search channels and applications sent out, is negatively correlated with an individual's happiness, in the sense that happier job seekers exert less job search effort.

The third strand of studies using the IZA ED Survey addresses different questions within the literature concerning the economics of migration. Besides information on job search behavior, the dataset includes detailed information on the migration and social background of individuals and their parents, language

skills, religious affiliation and ethnic identity. Using this information, Constant et al. (2011a) investigate the extent to which the native-migrant gap in the labor market (migrants face lower employment probabilities and earnings) can be explained by ethnic identity and social integration. Applying a recently developed concept to differentiate between groups of migrants in terms of ethnic identity, the so-called *ethnosizer* (developed by Constant et al., 2009), the authors find that ethnic identity plays an important role in explaining differences in employment outcomes between natives and migrants. The lower employment rates among less integrated migrants can be attributed to lower search effort and relatively high reservation wages.

Constant et al. (2010) address the question of why the native-migrant distance in terms of economic outcomes persists over migrant generations despite second generation migrants achieving higher educational outcomes than their parents. In fact, they test the hypothesis of whether second generation migrants (born in Germany) have higher reservation wages than first generation migrants (not born in Germany), given that the former tend to orientate towards the wage level in the host country while the latter refer to their country of origin (where wages are on average lower than in Germany). Indeed, they find higher reservation wages for second generation migrants, which might explain the persistence of the native-migrant gap in economic outcomes, although second generation migrant catch up in terms of educational attainment.

Constant et al. (2011b) extend the analysis of second generation migrants and compare them to natives in order to understand the persistence of the native-migrant gap. They find considerable differences in terms of attitudes and risk preferences, which however, do not explain lower employment probabilities among second generation migrants.

These existing studies illustrate the high potential of the IZA ED Survey for empirical research. They demonstrate as well that the range of potential research questions which can be addressed by the data is broad. However, the fact that the data have been collected by means of surveys and the focus on (initially) unemployed individuals provide natural restrictions to applications. Thus, the addressable research questions need to be focused on issues related to individual employment histories which start with registered unemployment. Research questions dealing, for example, with on-the-job search are not in the scope of the data. Two further restrictions which need to be taken into account are the non-negligible attrition (see section 3.3), in case researchers want to address dynamic questions, and potential measurement noise in survey responses on behavioral questions like reservation wages or personality traits etc.

Overall, however, it can be stated that the variety of information included in this survey allows researchers to contribute new insights to many different issues within the field of labor economics.

5. **Summary and Outlook**

This paper introduces the IZA ED Survey, which has been created to overcome data limitations in empirical labor research, particularly to provide more evidence about *how* successful job search and ALMP interventions operate. Beyond this aim, this panel survey can be used to study many issues within labor economics that set high demands on data richness. The new Scientific Use Files provided by the International Data Service Center of IZA cover a large and representative population of around 18,000 unemployed individuals who entered unemployment insurance in Germany between May 2007 and June 2008. The individuals were repeatedly interviewed over four waves in order that their labor market trajectories can be observed up to three years after unemployment entry. This large sample of unemployed individuals allows for more detailed and heterogeneity analyses (of subgroups, etc.) than a usual general-interest panel survey.

The core advantage of the IZA ED Survey is reflected in the *combination* of several types of crucial information within one data set: It provides very rich information on job search behavior, personal attitudes, traits, perceptions and characteristics as well as concerning the social and cultural environment of the surveyed individuals, including ethnicity and a migration background. This is combined with longitudinal data that track the individual pathways with respect to labor market activities and outcomes, as well as ALMP participations. Therefore, this data collection allows designing detailed studies regarding the

interplay of personal (search) behavior and attitudes, labor market outcomes and labor market policies.

Accordingly, the goal of the provision of the IZA ED Survey to the scientific community is to inspire more research about the mentioned interplay. Some potential future lines of research based on the IZA ED Survey could include the analysis of dynamics of some of the aforementioned aspects, as well as their impact on labor market outcomes. Evaluations of labor market policies can be enriched by the study of these aspects, in order to provide more empirical evidence on *how* ALMP needs to be designed in order to be successful. Moreover, potential research questions can go far beyond these topics. For instance, getting to know more about the search behavior of different subgroups of the population (different ages, different cultural backgrounds, etc.) can be instructive for future policy design. More generally, the IZA ED Survey provides a collection of data which allow for potentially innovative empirical research that *combines* issues of different economic subfields, like e.g. behavioral economics, unemployment insurance and welfare system design, education, migration and public economics.

Finally, the construction of the IZA ED Survey was part of a broader project aimed at creating a new data base to analyze social and labor policies (see Caliendo et al., 2011a for details). Thereby, the main feature is that the survey data, as presented here, can be merged with individuals' administrative data as provided by the IAB. The administrative data contain daily information on

individuals' time spent in employment, unemployment and participation in ALMP programs, including wages and benefits. Merging the survey with administrative data has the advantage that the variety of information included in the survey is enriched by highly reliable information on individuals' labor market activities and earnings, which are observable for a period that is much longer than the survey window (covering from 1975 until present). However, the administrative data are subject to German data security legislation, which prevents public access to the merged dataset. Therefore, we cannot yet provide the administrative information with the Scientific Use Files of the IZA ED Survey, although we are currently working – together with the IAB – on a solution to provide user access to the merged dataset in the future.

References

Akgüc, M., C. Giulietti, K. F. Zimmermann (2013): The RUMiC Longitudinal Survey: Fostering Research on Labor Markets in China. Working Paper, IZA Bonn.

Van den Berg, G., A. Bergemann, M. Caliendo (2009): The Effect of Active Labor Market Programs on Not-Yet Treated Unemployed Individuals. *Journal of the European Economic Association*, 7(2-3), 606-616.

Van den Berg, G., A. Bergemann, M. Caliendo, K. F. Zimmermann (2011): The Threat Effect of Participation in Active Labor Market Programs on Job Search Behavior of Migrants in Germany. *International Journal of Manpower*, 32 (7), 777-795.

Blossfeld, H.-P., H.-G. Roßbach, J. von Maurice (Eds.) (2011): Education as a Lifelong Process: The German National Educational Panel Study (NEPS). Wiesbaden. VS Verlag für Sozialwissenschaften.

Bonin, H., T. Dohmen, A. Falk, D. Huffman, U. Sunde (2007): Cross-Sectional Earnings Risk and Occupational Sorting: The Rolfe of Risk Attitudes. *Labour Economics*, 14 (6), 926-937.

Borghans, L., A. L. Duckworth, J. J. Heckman, B. ter Weel (2008): The Economics and Psychology of Personality Traits. *Journal of Human Resources*, 43 (4), 972-1059.

Caliendo, M., A. Falk, L. C. Kaiser, H. Schneider, A. Uhlendorff, G. J. van den Berg, K. F. Zimmermann (2011a): The IZA Evaluation Dataset: Towards Evidence-Based Labor Policy Making. *International Journal of Manpower*, 32 (7), 731-752.

Caliendo, M., W.-S. Lee (2013): Fat Chance! Obesity and the Transition from Unemployment to Employment. *Economics and Human Biology*, 11 (2), 121-133.

Caliendo, M., R. Schmidl, A. Uhlendorff (2011b): Social Networks, Job Search Methods and Reservation Wages: Evidence for Germany. *International Journal of Manpower*, 32 (7), 796-824.

Caliendo, M., A. Uhlendorff (2011): Determinanten des Suchverhaltens von Arbeitslosen: Ausgewählte Erkenntnisse basierend auf dem IZA Evaluationsdatensatz. *Journal for Labour Market Research*, 44 (1-2), 119-125.

Constant, A. F., L. Gataullina, K. F. Zimmermann (2009): Ethnosizing Immigrants. *Journal of Economic Behavior & Organization*, 69 (3), 274-287.

Constant, A. F., M. Kahanec, U. Rinne, K. F. Zimmermann (2011a): Ethnicity, Job Search and Labor Market Reintegration of the Unemployed. *International Journal of Manpower*, 32 (7), 753-776.

Constant, A. F., A. Krause, U. Rinne, K. F. Zimmermann (2011b): Economic Preferences and Attitudes of the Unemployed: Are Natives and Second Generation Migrants Alike?. *International Journal of Manpower*, 32 (7), 825-851.

Constant, A. F., A. Krause, U. Rinne, K. F. Zimmermann (2010): Reservation Wages of First and Second Generation Migrants. IZA Discussion Paper 5396, Bonn.

Constant, A. F., K. F. Zimmermann (2013): Migration, Ethnicity and Economic Integration. published in: A. Constant, K. F. Zimmermann (eds.), *International Handbook on the Economics of Migration*, Edward Elgar Publishing, Cheltenham, 13-36.

Constant, A. F., K. F. Zimmermann (2008): Measuring Ethnic Identity and Its Impact on Economic Behavior. *Journal of the European Economic Association*, 6 (2-3), 424-433.

Guillermina, J., D. S. Massey, M. R. Rosenzweig, J. P. Smith (2000): The New Immigrant Survey Pilot (NIS-P=: Overview and New Findings about U.S. Legal Immigrants at Admission. *Demography*, 37 (1), 127-138.

Huinink, J., J. Brüderl, B. Nauck, S. Walper, L. Castiglioni, M. Feldhaus (2011): Panel Analysis of Intimate Relationships and Family Dynamics (pairfam): Conceptual Framework and Design. *Journal of Family Research*, 23 (1), 77-100.

Kong, S. T. (2010): Rural-Urban Migration in China: Survey Design and Implementation. In: Meng, Xin and Manning, Chris (Eds.) with Shi, Li and Effendi, Tadjuddin. The Great Migration: Rural-Urban Migration in China and Indonesia, Edward Elgar Publ. Ltd.

Konle-Seidl, R., W. Eichhorst, M. Grienberger-Zingerle (2010): Activation Policies in Germany: From Status Protection to Basic Income Support. *German Policy*, 6, 59-100.

Krause, A. (2013): Don't Worry, Be Happy? Happiness and Reemployment. forthcoming in *Journal of Economic Behavior and Organization*, DOI: http://dx.doi.org/10.1016/j.jebo.2013.09.002.

OECD (2013): OECD Employment Outlook 2013. OECD Publishing.

Rinne, U., A. Uhlendorff, Z. Zhao (2012): Vouchers and Caseworkers in Training Programs for the Unemployed. forthcoming in *Empirical Economics*.

Solga, H., G. G. Wagner (2007): A Modern Statistical Infrastructure for Excellent Research and Policy Advice – Report on the German Council for Social and Economic Data During its First Period in Office (2004-2006). *Journal of Applied Social Science Studies*, 127 (2), 315-320.

---

[1] The German Federal Employment Agency reports an annual unemployment rate of 9.0% and 7.8% in 2007 and 2008, respectively.

[2] The IDSC is another initiative by IZA to improve data availability within labor economics. The IDSC is embedded into a larger recent initiative by the German Council for Social and Economic Data to create an infrastructure for data access and documentation in Germany (see Solga and Wagner, 2007). The idea is to establish a network of Research Data Centers and Data Service Centers in order to improve data access and transparency for the scientific community.

[3] The User Manual of the IZA ED Survey can be found at http://idsc.iza.org/iza-ed-survey.

[4] The time period was arbitrarily chosen but captures one complete year, so that seasonality in the labor market can be taken into account in empirical analyses.

[5] The survey was conducted by *infas,* the Institute for Applied Social Sciences, which is a private and independent market and social research institution in Bonn, Germany.

[6] We thank Martin Spiess (University Hamburg), Doris Hess and Reiner Gilberg (infas) for their advice on the non-response analysis.

[7] Despite the long interview duration, only 2-3% of the interview refusals reported that they refused to participate in the survey due to the interview duration (see user manual for a detailed analysis of interview refusals).

[8] The official labor market statistics („Arbeitsmarktberichte" of the German Federal Employment Agency) reports about 4.2 million unemployment entries (into SGB III) per year, on average for 2007 and 2008. Among those, bit more than 0.5 million entries into short-term training are registered, which corresponds to a participation rate of 13.2%. The figures for retraining are 5.6%, for public employment schemes 0.3%, for wage subsidies 3.3% and for start-up subsidies 2.9%. Note that these are stock figures, i.e. several participations per year and type of program can be registered. As a consequence, rates on short-term activities are higher in these statistics than in Table 7, and vice versa for longer-run activities.