Methodology
of Human Capital Study

2016-2023

Methodology of Human Capital Study
2016-2023

Author: Patrycja Antosz
English translation: GTC AMG Sp. z o.o.

The publication is also available on the website
www.parp.gov.pl

electronic document

Warszawa 2018

Graphic design:
www.ccpg.com.pl

Fundusze Europejskie
Wiedza Edukacja Rozwój

Rzeczpospolita Polska

PARP
Grupa PFR

Unia Europejska
Europejski Fundusz Społeczny

Table of contents

About the Human Capital Study (BKL)

Human Capital Study (BKL – pl Bilans Kapitalu Ludzkiego) carried out between 2016 and 2023 is a continuation of the first edition of studies conducted in 2010–2015. At that time, various administrative bodies at central and local level expressed the need for a wide range of information about human capital resources available on the Polish labour market. Over the course of five editions of studies carried out during this period a lot of valuable information was collected – it concerned inter alia the employers' demand for certain categories of workers, the recruitment methods, competence requirements which the candidates were expected to fulfil, the level of competence of Polish employees, as well as unemployed and economically inactive persons, pupils and students, development activities undertaken by Poles and the development activities offered by companies and training institutions. Information gathered over the course of the study was frequently used by various institutions at international (the European Commission, Cedefop), central (e.g. Ministry of Development, Ministry of Science and Higher Education) and local (Voivodeship Labour Offices and Marshall Offices) levels, as well as by academia. Indicators developed for the purpose of monitoring the labour market became a standard of sorts. All these factors contributed to the need to continue the study in a modified form allowing for measuring the phenomena related to the labour market and the changes it is undergoing even more precisely. It was decided that in the second edition of the BKL certain indicators will be expanded (e.g. those related to the development of the employee competences), and that the employers' and adults' surveys will have a stronger panel orientation. These modifications will allow for more precise tracking of changes in the most relevant indicators. Similarly to the first edition of HCS, the project is implemented by the Polish Agency for Enterprise Development in partnership with the Center for Evaluation and Analysis of Public Policies of the Jagiellonian University.

The main objective of this project is to analyse and monitor human capital resources existing on the Polish labour market and the factors contributing to its development. This objective may be achieved through the implementation of specific goals via applied surveys, ie.:

BEmployers' surveys, including:
- determining the demand for employees;
- determining the demand for competences;
- developing the human capital management strategies (awareness of the need to invest in human resources, applied human resources strategies).

Adults' surveys, including:
- determining the availability of employees in specific professions;
- determining the availability and the level of acquired competences;
- identifying undertaken development activities.

Surveys of institutions and companies providing development services, including:
- determining the characteristic features of institutions and companies providing development and training services;
- determining the characteristic features of the offered forms of expanding competences.

Table 1. Table 1 presents the different study modules within a specific time period.

	edition 2017	edition 2018	edition 2019	edition 2020	edition 2021	edition 2022
Employers Panel						
Employers Cross-sectional survey						
Adults Panel						
Adults Cross-sectional survey						
Development services Cross-sectional survey						

Structure of the report

The report contains a detailed description of methodological assumptions and the process of carrying out the Human Capital Study 2016–2023[1] in three modules:

1. employers' surveys;
2. adults' surveys; and
3. surveys of institutions and companies providing development/training services.

Descriptions of study modules consist of the following elements:

- main premises – description of the concept of the study together with a summary of the implementation of its subsequent editions;
- population – description of the surveyed population to which the study conclusions refer;
- sampling frame – description of the source of data on the basis of which enterprises to be included in the study were sampled;
- sampling scheme – general description of the sampling method;
- weights – description of analytical weights and conditions for their use.

Questionnaires used during field work are available on the project's website www.parp. gov.pl. Each year, after the completion of the field work,the main part of the report will be supplemented with information regarding current edition of the study.

---

[1]   As regards the detailed information about the way in which the study was carried out, the content of methodology reports submitted together with relevant data by the Public Opinion Research Center Foundation (adults' studies) and PBS Sp. z o.o. (the other modules) was used.

Employers' surveys

Main premises

Employers' surveys will be carried out between 2017 and 2022 in two forms: main surveys and interim surveys. Main surveys, organised every two years, are based on the following panels: a small enterprises panel (N = 1,000), a medium-sized enterprises panel (N = 500), a large enterprises panel (N = 500) and a cross-sectional sample (N = 1,500) composed in equal parts of small, medium and large enterprises. Interim surveys will include two panels: a large enterprises panel and a medium-sized enterprises panel.

Table 2. Assumed sample size in consecutive editions[2] of the survey.

Survey	edition 2017	edition 2018	edition 2019	edition 2020	edition 2021	edition 2022
Panel	2000	1000	2000	1000	2000	1000
Cross-sectional survey	1500		1500		1500	

Table 3. Summarised information concerning the implemen of employers' surveys.

a) Basic information

Fieldwork period	21.08.2017 –1.12.2017 (quality control 27.02.2018–06.04.2018)
Randomised sampling frame	Bisnode Polish Database
Randomised sampling frame	stratified, disproportionate
Data collection technique	3481 CAPI interviews (96%) + 163 CATI interviews (4%)
Number of interviews carried out	3644
Average interview duration	40 minutes
Net response rate	36%

---

[2]   Main editions of the survey have been marked in red, while interim editions in yellow.

b) Sampling

Sampling frame validity	06.2017
Stratifying variables	size, section of the PKD1
Size of the randomly selected sample (including the primary sample)	11667 (9723)
Number of clustersk	–
Cluster size	–

c) Fieldwork

Size of the sample used	11619
Sampling frame errors[3]	900
Refusals and other reasons for non-response	7075

Population

Just as in the previous editions of the Human Capital Study (carried out in 2010-2015), the present employers' survey covers enterprises employing at least one employee. In view of the subject of the survey, i.e. human capital, the sampling frame does not include:

1. Local units – all types of branches or local subsidiaries of enterprises. The sampling frame will only include "parent units", *i.e.* each operator will be uniquely identified by its REGON number.
2. Enterprises covered by section 94 of the 2007 PKD entitled "Activities of membership organisations".
3. Enterprises covered by sections:
   - A – agriculture, forestry, hunting and fishing;
   - O – public administration and defence; compulsory social security;
   - T – households as employers; goods- and services-producing activities of households for own use;
   - U – extraterritorial organisations and bodies.
4. Enterprises having the following specific legal form:
   - 48 – foundations;
   - 50 – the Catholic Church;
   - 51 – other churches and religious associations;
   - 55 – associations;

---

[3]   Described in detail in the part dedicated to sampling.

- 60 – social organisations not listed separately;
- 70 – political parties;
- 72 – trade unions;
- 73 – employers' organisations;
- 76 – economic and professional self-government;
- 85 – homeowner associations;
- 90 – associations of agricultural producers groups.

5. Self-employed persons.
6. Enterprises, for which no contact details necessary for carrying out the survey are available in the sampling frame

Sampling frame

The randomised sampling frame consists of enterprises listed in the Polish Database kept by Bisnode. This database has been compiled based on the data contained in the National Official Business Register (so-called REGON). In addition, it has been verified in terms of data validity and supplemented with contact information available online.

Sampling scheme

Randomly selected sample size shall reflect the assumed net response rate (36%), sampling frame error and reserve contacts. Based on previous experience, it has been assumed that the sampling frame error combined with reserve contacts shall correspond to 16.7% of the randomly selected sample.

The following formula is used to calculate the size of the randomly selected sample in the e edition of the survey (N_e^l):

$$N_e^l = 1{,}2 \ (N_e^z \ / \ 0{,}36)$$

where:

1,2 – adjustment for the sampling frame error and reserve contacts (1/(1-0.167)),

N_e^z – assumed size of the implemented sample in the e edition of the survey,

0,36 – assumed net response rate.

The analysed sample is divided into the base part and the reserve part. The reserve sample consists of contacts used to replace sampling frame errors, which include:

- enterprise's non-compliance with the survey criteria – the enterprise does not fulfil the criteria related to the number of employees employed or legal forms or PKD sections selected to be included in the survey;
- wrong contact details – correct contact details cannot be established based on generally available sources;
- non-existent enterprises – the enterprise has never existed or is currently either suspended or in liquidation;
- duplicate record.

Main surveys

The sample in the main surveys shall be randomly selected based on disproportionate stratification, taking into account the size class of a given enterprise (3 classes of enterprises):
- micro- and small (employing between 2 and 49 employees);
- medium-sized (employing between 50 and 249 employees);
- large enterprises (employing 250 or more employees);
- and the groups covered by relevant sections of the PKD, represented in the sample in equal numbers (6 categories):
- Manufacturing and mining (sections B, C, D, E);
- Construction and transport (sections F, H);
- Trade, accommodation, food and beverage service activities, support services (sections G, I, N);
- Specialised services (sections J, K, L, M, R, S);
- Education (section P);
- Human health and social work activities (section Q).

Panel sample

A part of the employers' survey will be carried out as a panel. Medium-sized (N = 500) and large (N = 500) enterprises will be surveyed within the panel annually, *i.e.* both under the main surveys as well as under interim surveys. In contrast, a sample consisting of 1,000 small enterprises employing less than 50 employees will be surveyed every two years, exclusively under main surveys.

Cross-sectional sample

The cross-sectional sample used for the purposes of carrying out main surveys (N = 1,500) will be divided into groups of equal size based on the size class of a particular enterprise. Table 4 provides accurate information about the sample sizes assumed for the individual enterprises size classes under subsequent editions of the survey.

Interim surveys

Interim surveys carried out in two-year intervals, *i.e.* between the main editions of the survey, will only cover the panel of medium-sized and large enterprises due to the low dynamics of the phenomena surveyed in micro- and small enterprises.

Respondent

In each of the randomly selected enterprises, the respondent is a person which is most familiar with the issues related to employment, required skills and professional skills development (training). Depending on the size and structure of enterprises, the respondent does not necessarily have to be the contact person.

Weights

1. **trunc_dweight_smpl** – a weight reflecting the structure of the population determined on the basis of stratifying variables used at sampling stage, within which observations representative for an extremely high number of enterprises receive a weight equal to one. The so-called Tukey's upper hinge, *i.e.* a value equal to the upper quartile plus the double of the interquartile range, has been adopted as the extremity criterion. This weight should be used when carrying out standard analyses. The weights are trimmed within particular size classes of enterprises.
2. **trunc_dweight_pop** – a weight with a trunc_dweight_smpl structure. After it is applied, the sample size is added up to the population size. This weight should be used in cases where the analysis is conducted on variables which can be meaningfully interpreted in population-related categories, *e.g.* to estimate the number of employees needed in particular sectors of the economy.
3. **dweight_smpl** – an analytical weight consisting of a combination of weights associated with a given sampling scheme and non-response weights, which provides the sample

with a population structure; it is nevertheless added to the sample size. This weight is used to control the effect of trimming carried out for the purposes of applying the above-described weights.

4. **dweight_pop** – weight with a dweight_smpl structure which has the same applications as dweight_smpl. After it is applied, the sample size is added up to the population size.

Table 4. Assumed implemented samples sizes in consecutive editions of the survey in terms of panel participation.

a) Main surveys: editions 2017; 2019; 2021

Total (edition) = 3500

enterprise size	2-49	50-249	250+
Cross-sectional survey (main survey)	1000	500	500
Total (edition+size)	500	500	500
Total (edition+size)	1500	1000	1000

b) Interim surveys: editions 2018; 2020; 2022

Total (edition) = 1000

enterprise size	2-49	50-249	250+
Panel (interim surveys)	0	500	500
Cross-sectional survey (main survey)	0	0	0
Total (edition+size)	0	500	500

c) Total

Total (edition) = 13500

enterprise size	2-49	50-249	250+
Panel (interim surveys)	3000	3000	3000
Cross-sectional survey (main survey)	1500	1500	1500
Total (edition+size)	4500	4500	4500

Adults' surveys

Main premises

Adults' surveys will be carried out between 2017 and 2023 in two forms: main surveys and interim surveys. Both forms of the study will be carried out three times every two years. Each edition of the main survey will consist of 4,000 interviews divided into panel sample (N = 1,500) and a cross-sectional sample (N = 2,500). On the other hand, interim surveys consist of the panel sample (N = 1,500) only.

Table 5. Assumed sample size in consecutive editions[4] of the survey.

Survey	edition 2017	edition 2018	edition 2019	edition 2020	edition 2021	edition 2022
Panel	1500	1000	1500	1500	1500	1500
Cross-sectional survey	2500		2500		2500	

Table 6. Summary of the implementation of adults' surveys.

a) Basic information

Fieldwork period	21.08.2017–07.11.2017
Randomised sampling frame	PESEL
Sampling scheme	stratified, proportionate
Data collection technique	CAPI
Number of interviews carried out	4056
Average interview duration	37 minut
Net response rate	56%

---

4   Main editions of the survey have been marked in red, while interim editions in yellow.

b) Sampling

Sampling frame validity	07.2017
Stratifying variables	subregion, size of the locality, gender + age
Size of the randomly selected sample (including the primary sample)	11609 (7144)
Number of clusters	893
Cluster size	13

c) Fieldwork

Size of the sample used	11319
Sampling frame errors[5]	3552
Refusals and other reasons for non-response	3711

Population

The surveyed population consists of Polish citizens aged 18–69 residing in Poland.

Sampling frame

Contact details were taken from the Universal Electronic System for Registration of the Population (PESEL) managed by the Ministry of Digital Affairs.

Sampling scheme

Main surveys

Sampling is carried out in a stratified and proportionate manner, taking into account the Central Statistical Office's subregion, the given locality class size and gender combined with age. The sampling process is based on territorial stratification carried out in line with the division into Central Statistical Office's subregions (NTS-3, *i.e.* 73 subregions on the territory of the entire country, several in each voivodeship[6]) and localities class sizes determined in accordance with Central Statistical Office's criteria (9 classes):

---

[5] Described in detail in the part dedicated to sampling.

[6] A division applicable since 1 January 2018 has been utilised.

- villages (non-urbanised areas);
- cities (urbanised areas), divided into the following size classes:
  - up to 9,999 inhabitants;
  - 10,000–19,999 inhabitants;
  - 20,000–49,999 inhabitants;
  - 50,000–99,999 inhabitants;
  - 100,000–199,999 inhabitants;
  - 200,000–499,999 inhabitants;
  - 500,000–999,999 inhabitants;
  - Warsaw.

When developing the assumptions for the 2017 edition of the study, 310 strata were identified as a result of combining the two variables. Under each of the localities size classes, primary sampling units (PSU) – *i.e.* municipalities or parts of municipalities treated as separate "administrative units" – were selected. In the case of urban-rural municipality, the rural part of the municipality (belonging to class 1 – villages) will be treated separately from its urban (city) part belonging to the appropriate city class based on the actual number of inhabitants. Therefore, PSUs include rural municipalities, rural parts of urban-rural municipalities, urban parts of urban-rural municipalities and urban municipalities. In the first phase of the procedure, PSUs are sampled with replacement proportionally to the size of the population aged 18–69, assuming that a fieldwork cluster consists of thirteen people, and that each stratum is represented by at least one cluster.

In the second phase of sampling procedure, within each PSU persons belonging to the surveyed population are subjected to simple sampling. Stratification of persons sampled during the second phase takes into account the distribution by gender – age cohorts for the different localities size classes. The following age ranges have been established as age cohorts:

- women aged 18–29;
- women aged 30–39;
- women aged 40–49;
- women aged 50–69;
- men aged 18–29;
- men aged 30–39;
- men aged 40–49;
- men aged 50–69.

The established sample size shall reflect the assumed net response rate (56%), sampling frame error and reserve contacts. Based on previous experiences, it has been assumed that the sampling frame error combined with reserve contacts shall correspond to about 38.5% of the established sample size.

The following formula is used to calculate the size of the randomly selected sample in the *e* edition of the survey (N_e^l):

$$N_e^l = 1{,}625 \ (N_e^z / 0{,}56)$$

where:

1,625 – adjustment for the sampling frame error and reserve contacts (1/(1-0.3846)),

N_e^z – assumed size of the implemented sample in the *e* edition of the survey,

0,56 – assumed net response rate.

The size of the randomly selected sample is rounded to the nearest multiple of the number of contacts within a given fieldwork cluster.

Surveyed clusters are divided into the base part and the reserve part. Base sample constitutes an 8-person portion of the fieldwork cluster consisting of contacts which make it possible to carry out the survey in line with the adopted response rate indicator. Reserve sample constitutes a 5-person portion of the fieldwork cluster consisting of contacts used to replace sampling frame errors, which include:

- permanent emigration of the respondent – the person selected to take part in the survey is currently abroad and there is no information about his/her intention to return;
- the respondent could not be found at his/her place of residence – the person selected to take part in the survey (1) does not reside at the indicated address and is not known to the people residing under this address or (2) resides in a collective household (penal institution, hospital, etc.);
- incorrect address – the address indicated in the sampling frame (1) does not exist; (2) is an address of an uninhabited building; or (3) is not an address of a private residence;
- incorrect data about the respondent – the person selected to take part in the survey (1) has not reached the age of 18; (2) does not speak Polish; or (3) has died.

Interim/panel surveys

Constituent panel sub-sample consisting of 1,500 people was randomly selected from the persons who had taken part in the first edition of the study and had agreed to further

contact. In addition, after the 2017 edition of the study the randomised sampling frame included only people aged 18–63. This remains in line with the assumption that in the last year of the study respondents taking part in the panel will still belong to the surveyed population, i.e. will not be older than 69 years of age. The structure of the panel, similarly to the structure of cross-sectional surveys, reflects the structure of the sample in main surveys. Respondents randomly selected to take part in the panel, who – for various reasons – do not take part in subsequent editions of the study are substituted by other persons belonging to the same stratum. In order to gather the largest possible amount of data for the purpose of tracing surveyed persons, people substituting the surveyed persons are first randomly selected from persons who took part in previous editions of the survey. This also means that, due to the short duration of the panel, people aged 18 are not randomly selected to participate in the panel in the subsequent years.

Weights

Four types of weights have been developed:

1. **trunc_dweight_smpl** – a weight reflecting the structure of the population determined on the basis of stratifying variables used at sampling stage[7], within which respondents representing extremely high number of persons receive a weight equal to one. The so-called Tukey's upper hinge, *i.e.* a value equal to the upper quartile plus the double of the interquartile range, has been adopted as the extremity criterion. This weight should be used when carrying out standard analyses.

2. **trunc_dweight_pop** – a weight with a trunc_dweight_smpl structure. After it is applied, the sample size is added up to the population size. This weight should be used in cases where the analysis is conducted on variables which can be meaningfully interpreted in population-related categories, *e.g.* to estimate the number of people in Poland who were employed on the basis of civil law contracts over the course of the last 12 months.

3. **dweight_smpl** – an analytical weight consisting of a combination of weights associated with a given sampling scheme and non-response weights, which provides the sample

---

[7] Due to the relatively small sample size, the weighting scheme does not ideally reflect the sampling scheme in the case of division into subregions and division by size classes of localities. For the purposes of weighting, a division into 7 macroregions (in line with NTS 1) and into 4 size classes of localities, i.e. (1) villages, (2) small cities (<20 thousand inhabitants), (3) medium-sized cities (20–99 thousand inhabitants) and (4) large cities (>100 thousand inhabitants), was adopted. In the case of division by gender and age, the weighting scheme corresponds to the sampling scheme.

with a population structure; it is nevertheless added to the sample size. This weight is used to control the effect of trimming carried out for the purposes of applying the above-described weights.

4. **dweight_pop** – weight with a dweight_smpl structure which has the same applications as dweight_smpl. After it is applied, the sample size is added up to the population size.

Surveys of institutions and companies providing development services

Main premises

Surveys will be carried out in the form of main surveys three times between 2017 and 2023 in two-year intervals. Each edition will consist of 1,000 interviews.

Table 7. Assumed sample size in consecutive editions of the survey.

Survey	edition 2017	edition 2018	edition 2019	edition 2020	edition 2021	edition 2022
Cross-sectional sample	1000		1000		1000	

Table 8. Summary of the surveys of institutions and companies providing development services.

a) Basic information

Fieldwork period	24.08.2017–07.12.2017 (quality control 26.02.2018–18.04.2018)
Randomised sampling frame	own base
Sampling scheme	stratified, disproportionate
Data collection technique	859 interviews CAPI (86%) + 145 interviews CATI (14%)
Number of interviews carried out	1004
Average interview duration	43 minut
Net response rate	31%

b) Sampling

Sampling frame validity	08.2017
Stratifying variables	type of institution
Size of the randomly selected sample (including the primary sample)	3388 (3226)
Number of clusters	–
Cluster size	–

c) Fieldwork

Size of the sample used	3363
Sampling frame errors[8]	252
Refusals and other reasons for non-response	2107

Population

Under the Human Capital Study 2016–2023 a decision was taken to establish a more precise, and therefore narrower, definition of an enterprise identified as an institution or company providing development/training services. For the purposes of this study, such an institution or company shall be regarded as an entities providing development services (*i.e.* professional courses aiming at improving qualifications and professional skills courses, certified development programmes, group training sessions and exercises, individual training sessions and exercises, conferences, seminars, schools for adults, post-graduate studies, lectures, read-outs, talks, consulting services, advisory services, coaching, mentoring, workshops, events organised to provide instructions) within the scope of non-formal and formal education which goes beyond the standard education cycle[9], if:

1. it provides development services for adults (people aged 18+) on the territory of Poland;
2. it is an independent institution/company, including sole traders and excluding those enterprises who provide development services exclusively to a single institutional client;

---

[8]   Described in detail in the part dedicated to sampling.

[9]   We consider that the standard education cycle consists of educating pupils and students aged 6–24 and who typically complete the relevant education stages without interruptions in a stationary mode at primary, lower-secondary, upper-secondary and higher level (1st, 2nd and 3rd degree).

3. it is the head office of a multi-branch institution/company or is one of the branches of a larger institution/company and has a separate legal personality, *i.e.* its REGON number is different from the REGON number of the head/central office;

4. it carries out its activities under a franchise;

5. the nature of development services provided by it does not consist exclusively in services from the list presented below:

   a. courses in the field of occupational health and safety (OHS);

   b. courses in the field of fire safety;

   c. courses leading to attainment of driving competences (driving schools);

   d. hobby activities (e.g. sports activities, recreational activities);

   e. activities of incidental nature (e.g. birthing schools);

   f. cultural activities (e.g. services provided by libraries, cultural centres);

   g. development services for agricultural sector (farmers, not entrepreneurs from rural regions), e.g. Agricultural Advisory Centres;

   h. outsourcing services (e.g. provided by accountancy offices, legal council offices);

6. it is not a part of the Public Employment Services (*e.g.* Labour Offices);

7. generated turnover of at least PLN 50 thousand from the provision of development services (excluding services listed under point 5) for adults in line with the data from the calendar year preceding the survey.

Due to the fact that the survey focuses mainly on stable entities having a long-standing tradition of providing training services, the first edition of the study will focus on entities that have been operating on the development services market for at least three years. In subsequent editions of the study, population will be supplemented with databases containing data of units operating on the market for a period shorter than 3 years, though their share in the randomly selected sample cannot exceed 10% (depending on the available data, the share of such enterprises in particular strata may be different).

Sampling frame

The existing databases served as the basis for establishing the complete randomised sampling frame; these databases included:

- data gathered within the Educational Information System (hereinafter: SIO);
- data gathered within the POL-ON system (hereinafter: POL-ON);
- data gathered within the Development Services Database (hereinafter: BUR);

- data gathered within the Registry of Training Institutions (hereinafter: RIS);
- data gathered for the purposes of conducting the survey of training institutions and companied under the Human Capital Study 2010–2015 (hereinafter: CEAPP).

The database constituting the potential sampling frame that consists of 30,350 enterprises has been adequately supplemented and verified prior to the start of activities related to random selection. This procedure lasted from June to August 2017 and consisted of two stages conducted in parallel:

1. Verification whether a given respondent is a member of the surveyed population by phone. The results of verification by phone are presented in table 9.

Table 9. Results of verification carried out to establish whether a given spondent is a member of the surveyed population.

Verification	Source: SIO	Source: POL-ON	Source: BUR	Source: CEAPP	Source: RIS	Total
negative verification	378	41	521	1561	1364	3865
positive verification	217	34	141	631	959	1982
contact established despite the lack of verification	1924	252	1386	7487	5251	16300
failure to establish contact with an enterprise	392	54	296	6350	1111	8203
Total	2911	381	2344	16029	8685	30350

2. Supplementing the data contained in the different databases and verifying validity of such data via the Internet.

   As a result of this procedure, data necessary to conduct the survey were supplemented and verified; such data included:

- enterprise's address,
- phone number,
- REGON number.

Thanks to these supplementations, outdated entries concerning relevant enterprises and entries duplicated in several databases were removed (5,864 enterprises).

In the subsequent editions of the study, database containing name, address and phone number of relevant entities will be verified and supplemented, taking into account changes on the market of institutions and companies providing development services (certain institutions leaving the market, other institutions entering the market, change of seat, addresses etc.).

Every year, a sample of institutions and companies providing development services will be randomly divided into base sample and reserve sample. The size of the base sample will be calculated in such a manner as to make it possible to carry out the survey taking into account the rates of return imposed by the Contractor.

Sampling scheme

Due to the fact that one of the objectives of the study was to compare different types of enterprises, the final decision on whether a given enterprise belongs to a particular sampling frame was made based on this enterprise's affiliation with a particular randomised sampling stratum. Stratification was carried out exclusively based on the type of enterprise, in line with its affiliation presented in the primary source of data. Four types of enterprises were identified:

1. enterprises included in SIO database – CKU, CKP, ODZ, post lower-secondary schools;
2. enterprises included in POL-ON database – universities;
3. enterprises included in BUR – private enterprises registered in the Database of Development Services (BUR);
4. enterprises included in CEAPP or RIS – private enterprises not registered in the Database of Development Services.

Table 10 presents individual strata and their verification statuses, both in the randomised sampling frame and in the randomly selected sample.

Table 10. Randomised sampling frame and randomly selected sample by results of phone verification.

Verification	SIO	POL-ON	BUR	CEAPP	RIS	Total
Sampling frame: positive verification	208	137	137	608	955	1942
Sampling frame: contact established despite the lack of verification	1618	1288	1288	0	0	3121
Sampling frame: failure to establish contact with an enterprise	148	170	170	0	0	370
Sampling frame: Total	**1974**	**1595**	**1595**	**1563**	**5433**	
Sample: positive verification	208	137	137	413	616	1408
Sample: contact established despite the lack of verification	690	735	735	0	0	1640
Sample: failure to establish contact with an enterprise	131	157	157	0	0	340
Sample: Total	**1029**	**1029**	**1029**	**1029**	**3388**	

Randomly selected sample size shall reflect the assumed net response rate (31%), sampling frame error and reserve contacts. In light of earlier verification-related work, it has been assumed that the sampling frame error combined with reserve contacts shall correspond to 5% of the randomly selected sample.

The following formula is used to calculate the size of the randomly selected sample in the *e* edition of the survey (N_e^l):

$$N_e^l = 1{,}05 \ (N_e^z \, / \, 0{,}31)$$

where:

1,05 – adjustment for the sampling frame error and reserve contacts,

N_e^z – assumed size of the implemented sample in the *e* edition of the survey,

0,31 – assumed net response rate.

The analysed sample is divided into the base part and the reserve part. The reserve sample consists of contacts used to replace sampling frame errors, which include:

- enterprise's non-compliance with the survey criteria – the enterprise does not fulfil the criteria related to the number of employees employed and legal forms or PKD sections selected to be included in the survey;
- wrong contact details – correct contact details cannot be established based on generally available sources;
- non-existent enterprises – the enterprise has never existed or is currently either suspended or in liquidation;

Weights

Due to the fact that during sampling and weighting of data the same single stratifying variable was used, and thus all observations within a given stratum had the same analytical weights, two types of weights were developed:

1. **dweight_smpl** – an analytical weight consisting of a combination of weights associated with a given sampling scheme and non-response weights, which provides the sample with a population structure; it is nevertheless added to the sample size. This weight should be used when carrying out standard analyses.
2. **dweight_pop** – weight with a dweight_smpl structure which has the same applications as dweight_smpl. After it is applied, the sample size is added up to the population size. This

weight should be used in cases where the analysis is conducted on variables which can be meaningfully interpreted in population-related categories, e.g. to estimate the number of people in Poland who participated in various kinds of training over the course of the last 12 months

PARP
Grupa PFR

www.parp.gov.pl