

European Agency for Safety and Health at Work

ESENER-3: Quality Report

Table of Contents

Figures and tables	3
1 Foreword.....	5
2 Introduction to the statistical process and its outputs.....	6
2.1 General survey concept.....	6
2.2 Structure of the questionnaire	6
2.3 Data outputs from the statistical process.....	7
3 Relevance.....	8
4 Accuracy	10
4.1 Errors of non-observation	11
4.2 Errors of observation (measurement errors)	30
4.3 Processing errors.....	39
5 Timeliness and punctuality	41
6 Coherence and comparability.....	45
6.1 Development of the survey concept	45
6.2 Conceptualisation and organisation of the survey.....	49
6.3 Usage of harmonized national questionnaire versions.....	50
6.4 Comparability of the data – indications from the data-set	53
7 Respondent burden	55
8 Confidentiality, transparency and security.....	57
9 Annexes	58
10 Literature.....	76

Figures and tables

Figure 1: Summary of the questionnaire development process.....	31
Figure 2: Summary of the translation process.....	53
Table 1: List of abbreviations.....	4
Table 2: Establishments in the education and the health sector, by ownership type	16
Table 3: Definition of response, cooperation and other rates	18
Table 4: Cooperation and response rates by country, different calculation variants	19
Table 5: Share of item non-response per question (unweighted results)	23
Table 6: Sampling rates, by country	27
Table 7: Standard error (s.e.) and confidence interval (c.i.) at different levels of disaggregation.....	29
Table 8: Number and share of interviews done with external OSH consultants, by country	34
Table 9: Number and share of interviews done with external OSH consultants, by size.....	34
Table 10: Summary or multi-variate regression analyses on mode effects in ESENER-3	37
Table 11: Mode effects in the Czech subsample	38
Table 12: Mode effects in the subsamples for DK, ES, LU and NL (countries <100 online int.)	39
Table 13: Time periods for key project steps (plan, reality and comments).....	42
Table 14: Number and share of multi-site organisations and of subsidiaries, by country, unweighted	47
Table 15: Share of multi-site organisations and subsidiaries, by country, establ-prop. weighted	48
Table 16: List of additional questions asked in boost countries only	51
Table 17: Respondent burden in terms of interviewing time, by country	55
Table 18: Model 1: Dependent variable Q250 (Risk Assessments).....	59
Table 19: Model 2: Dependent variable Q356_1: Employee training the workstation	61
Table 20: Model 3: Dependent variable Q355_3: Training on prevention of PSR (Q355_3).....	63
Table 21: Model 4: Dependent variable Q311: Discussion of OSH impacts of new technologies.....	65
Table 22: Model 5: Dependent variable Q300: Action plan to prevent work related stress	67
Table 23: Model 6: Dependent variable Q159: Development of sickness absence.....	69
Table 24: ESENER-3 outcome codes by country, numerical.....	73

Table 1: List of abbreviations

Abbreviation	Full name/explanation
b2b survey	B usiness t o b usiness survey, i.e. survey among organisations (be it at the company/enterprise or at the establishment/local unit level)
CATI	C omputer A ssisted T elephone I nterviewing
CAWI	C omputer A ssisted W eb I nterviewing (online interviews)
DK	D on't K now (Answer category in the questionnaire)
ESENER	E uropean S urvey of E nterprises on N ew and E merging R isks
ESS	E uropean S tatistical S ystem
EU-OSHA	European A gency f or S afety and H ealth at W ork
LFS	L abour F orce S urvey
NA	N o A nswer (answer category in the questionnaire)
OSH	O ccupational S afety and H ealth (also called "Health and Safety at Work")
PSR	P sycho- s ocial risks (at work)
SBR	S tructural B usiness R egister (register of business addresses compiled by each EU member state under coordination of EUROSTAT)
SBS	S tructural B usiness S tatistics (statistics compiled by the national statistical offices on base of the SBR)
TP	T arget P erson
TripleC	Kantar coordination centre for multi-country telephone surveys

Country abbreviations (in alphabetical order)			
AT	Austria	IT	Italy
BE	Belgium	LT	Lithuania
BG	Bulgaria	LU	Luxembourg
CH	Switzerland	LV	Latvia
CY	Cyprus	MK	North Macedonia
CZ	Czech Republic	MT	Malta
DE	Germany	NL	Netherlands
DK	Denmark	NO	Norway
EE	Estonia	PL	Poland
EL	Greece	PT	Portugal
ES	Spain	RO	Romania
FI	Finland	RS	Serbia
FR	France	SE	Sweden
HR	Croatia	SI	Slovenia
HU	Hungary	SK	Slovakia
IE	Ireland	UK	United Kingdom
IS	Iceland		

1 Foreword

This is the quality report of ESENER-3, the “3rd European Survey of Enterprises on New and Emerging Risks”, conducted in 2019 on behalf of the European Agency for Safety and Health at Work (EU-OSHA) by the survey research institute Kantar GmbH, Public Division Munich and the Kantar network of field institutes.

The report closely follows the Eurostat guidelines for quality reports on statistical data, as documented in the “ESS¹ Standard for Quality Reports” (Eurostat Luxembourg, 2009a) and the more comprehensive “ESS Handbook for Quality Reports” (Eurostat Luxembourg, 2009b). These guidelines were developed for all kinds of statistical outputs and are mainly targeted at the national statistical offices within the EU. But a large part of the quality criteria around which these guidelines are structured apply also to sample surveys carried out by private survey research institutions.

Quality criteria dealt with in the ESS standards that are not relevant for this type of statistical output or which are not relevant at the time of writing this report are omitted. This is for example the case for “Accessibility and Clarity”, which, as discussed in the “ESS Standard for Quality Reports”, refer to the possibilities (outside EU-OSHA) to access the collected data and to the supporting information made available. As done already for the two previous waves, EU-OSHA will make the ESENER-3 data-set and documentation accessible at a later stage². But the ease of access and the clarity of the survey information made available for external researchers cannot yet be fully assessed, particularly not by those in charge of elaborating the technical reporting.

The quality report is not the only document related to the methodology of ESENER-3. In addition, the following reports are available:

- Technical Report, documenting the sampling and other methodological issues, fieldwork organisation, fieldwork outcomes and further information relevant for users of the ESENER-3 data-set.
- Translation Report, documenting the process of the elaboration of national questionnaire versions in detail.
- Pre-test Report, documenting the questionnaire, fieldwork and findings of the face-to-face cognitive pre-test carried out as first empirical test step for ESENER-3.
- Pilot Survey Report, documenting the set-up and outcomes of the CATI pilot interviewing done as last stage in the testing of the survey instrument in all 33 countries.

Considerable overlaps between the Quality Report and the Technical Report are unavoidable since the Technical Report was conceptualized as a document covering all issues that may be relevant for users of the data. The Quality Report partly analyses the same aspects, though from a different perspective. For several issues, reference will be made to the Technical Report for more details instead of repeating the respective passages from the Technical Report in detail here.

¹ European Statistical System

² The data-sets for ESENER-1 and ESENER-2 can already be accessed via the UK Data Archive in Essex (<https://www.data-archive.ac.uk>). EU-OSHA plans to make the ESENER-3 data-set available there towards the end of the year 2020. The data will additionally also be made available at the DAS, the Data Archive for Social Sciences located at the GESIS research institute in Mannheim, Germany (<https://www.gesis.org/en/institute/departments/data-archive-for-the-social-sciences>). As was the case already for ESENER-2, core results of the survey will additionally be placed on the EU-OSHA website, connected to a data visualization tool allowing e.g. to see results for individual countries and to compare them across countries, sectors and sizes: <https://osha.europa.eu/en/european-survey-enterprises-new-and-emerging-risks-esener>.

2 Introduction to the statistical process and its outputs

2.1 General survey concept

ESENER-3 is a cross-national survey instrument developed for direct collection of information on health and safety provisions and risk perceptions from respondents at the workplace level. ESENER-3 was carried out as a CATI survey, with an additional CAWI option for refusers of the telephone interview, aimed at reducing non-response (sequential mixed-mode design). The sampling unit and unit of enquiry were establishments (local units). Within the contacted units, the person who “knows best about health and safety in the establishment” was targeted.

Sampling was based on a probabilistic sampling procedure. The survey covered establishments from 5 employees onwards in almost all sectors of activity, with the sole exceptions of NACE Rev.2 sections “T” (Activities of households as employers) and “U” (Activities of extraterritorial organisations and bodies). It included 33 countries (the current EU-28 plus Iceland, North Macedonia, Norway, Serbia and Switzerland).

The coordination and overall responsibility for the project on part of the contractor was with Kantar GmbH, Public Division Munich. Fieldwork itself was coordinated centrally by the international TripleC CATI centre of Kantar with offices in Brussels and Prague. All interviews were carried out locally by Kantar institutes or cooperation partners within its international network. Addresses for the survey were drawn from business address registers of different origin. For the drawing of the gross sample and the steering of the net sample a multi-stratified random sampling procedure was applied. The disproportionalities of this sample design were afterwards corrected by weighting.

ESENER-3 is the third survey of its kind conducted in Europe, following ESENER-1, carried out in 2009, and ESENER-2, in 2014. All three survey waves generally cover the same topics but the questionnaire was subject to a major revision in ESENER-2 (2014). A significant proportion of the ESENER-2 questionnaire has remained unchanged in ESENER-3 (2019), enabling comparisons between the two last waves.

2.2 Structure of the questionnaire

The questionnaire developed for ESENER-2 consists of the following main parts:

- Section 1: Contact phase
Introduction of the survey, identification of the right target person within the establishment, provision of motivation letters and scheduling of interviews
- Section 2: Reminder and other call-backs
Technical section (recalls for reminders on online interviews and similar questions)
- Section 3: Establishment and sector screening
Screening of enterprise addresses on the existence of further local units, applicable only to the 17 countries with no establishment-level address register
- Section 4: Introductory questions
Background data on the establishment (sector, size, employee structure etc.) and on the respondent (function within the establishment)
- Section 5: Day-to-day OSH management I: OSH expertise and general policy
Mapping of expert support structures, general measures taken for health promotion and outcome indicators

- Section 6: (Traditional and new) health and safety risks in the establishment
Mapping of different types of health risks present due to the type of work to be performed or the way it is managed
- Section 7: Day-to-day OSH management Part II: Risk assessments
Implementation and design of workplace risk assessments respectively the reasons for not performing them
- Section 8: Psychosocial risks and digitalisation
Perception of different forms of psycho-social risks and measures taken to prevent these
- Section 9: Employee participation in health and safety issues
Mapping of institutional representation structures and participation practice
- Section 10: Questions specific to boost countries
- Section 11: Final background questions

2.3 Data outputs from the statistical process

The main output of the survey consists of a data-set with the answers of respondents to all survey questions. All questions except for one were closed questions. For the only open-ended question (Q110, asking for the main activity of the establishment), answers were coded and made available in a closed format (NACE Rev. 2 sector of activity).

The following data outputs have been made available to EU-OSHA:

- A labelled SPSS data-set, including the answer codes for all questions and different types of weights.
- A gross sample file with (fully anonymised) contact and outcome information for all addresses used for the survey.
- Establishment and employee proportionally weighted cross-tabulations showing the results of all questions by country.
- Establishment and employee proportionally weighted cross-tabulations showing the results for all 33 countries by size-class and by sector groups (not interrelated).

3 Relevance

In the “ESS Standard for Quality Reports”, relevance is defined as “the degree to which statistical outputs meet current and potential user needs. It depends on whether all the statistics that are needed are produced and the extent to which concepts used (definitions, classification etc.) reflect user needs” (Eurostat 2009a, p.7).

In the development phase of the questionnaire, various sources of feedback on the survey were taken into account in order to ensure the relevance of the survey outputs:

- Feedback of users at EU-OSHA that worked with the data of the preceding survey wave (ESENER-2).
- Feedback of users outside of EU-OSHA that used data from the preceding survey wave (ESENER-2).
- The outcome of the “Technical Assessment of the expansion of the Second European Survey of Enterprises on New and Emerging Risks (ESENER-2)” evaluating the second survey wave
- Wishes of the European Commission and other stakeholders for the inclusion of additional aspects in ESENER-3.
- Information needs of different departments within EU-OSHA.

Not all recommendations and wishes for amendments from the data users could finally be taken up in the ESENER-3 questionnaire and survey concept. Some recommendations for new questions were tested in the cognitive pre-test or the pilot but were finally discarded because pre-test results had raised doubts on the validity of the answers or because there was not enough room for all additional questions.

The extensive pre-testing phase with 36 cognitive pre-test interviews in three countries (DE, LV, NL) and a standardized pilot survey phase in all 33 countries were crucial steps in the efforts to enhance the relevance of the new questionnaire. Particularly in the cognitive pre-test interviews, respondents were given the chance to assess the relevance of the new questions for their establishment and to provide feedback on questions they considered as irrelevant or misleading. While this assessment was in general positive, a number of respondents from small establishments considered some of the aspects asked about in the questionnaire as not being relevant to them in view of the small size of their workplace. As consequence, in the main survey, questions capturing these aspects were partly filtered on the size of the establishment.

The two main topical amendments finally made to the questionnaire in order to enhance its relevance for researchers, politicians and practitioners are questions related to health and safety aspects of new technologies (Q310-Q312) and two rough indicators on the outcome of OSH measures in terms of sickness and absence rates (Q159 and Q160).

The wish to further enhance the relevance of the survey was also a main driver behind the decision to define a set of trend questions to be taken over in an identical form from the questionnaire of the preceding survey wave ESENER-2. The trend questions allow for an observation of developments over time, thus providing at least indications³ on whether national or international campaigns on specific

³ As ESENER is not a panel survey, it is not possible to clearly establish causal relationships between the introduction of certain measures (new legislation, campaigns on specific health and safety topics etc.) and changes in the health and safety practice at the establishment level.

issues (such as the improvement of workplace risk assessments or the raising of awareness for psycho-social risks at work) have had an impact.

As these measures show, the further enhancement of the relevance of ESENER for all potential data users was an important issue in the set-up of the new survey wave.

4 Accuracy

Definition:

“The accuracy of statistical outputs in the general statistical sense is the degree of closeness of estimates to the true values” (Eurostat 2009, p.8).

Any probability sample survey can be just an approximation to the real situation under investigation. It is inevitably based on estimations, trying to draw conclusions from the situation in some units to the situation in all units within the defined universe. The degree of accuracy to which a survey measures a phenomenon is influenced by various factors.

In the following, the main potential sources of errors inherent to sample surveys will be tackled and it will be analysed in how far ESENER-3 is affected by these potential sources of errors. Hereby, the classification of survey errors as introduced by Robert M. Groves will be used as guideline. In his book on survey errors, Groves (2004)⁴ introduces the following categorization for the possible sources of survey errors, and thus identifies the two major or core design pillars of a survey:

Errors of non-observation, i.e. because of not observing the whole target population (referred to in the following under the summarizing term “sampling errors”):

- Coverage (frame errors)
- Non-response
- Selection of a subset of the population⁵ (sampling errors)

Errors during observation (referred to in the following as „measurement errors”)

- Interviewer effects
- Respondent effects
- Effects of the survey instrument (questionnaire)
- Effects of the mode of data collection

As a third type of errors, **processing errors** are discussed in this chapter, distinguishing between two types of processing errors:

- Errors from data coding, editing and processing
- Weighting errors

⁴ Groves, Robert M. (2004): Survey Errors and Survey Costs. Hoboken, New Jersey: John Wiley & Sons.

⁵ Groves uses the term „sampling” for the selection of a subset of the population, thus addressing the implication that statistics are estimated on base of the randomly selected sample (see below „sampling error”).

4.1 Errors of non-observation

The first main type of errors identified by Groves is errors of non-observation, i.e. errors resulting from the situation that not all (types of) units that belong to the defined universe are properly observed. The main types of errors due to non-observation are frame errors, non-response errors and sampling errors. The occurrence of these errors in ESENER-3 will be analysed in the following chapters.

4.1.1 Coverage errors (frame errors)

Coverage errors are errors resulting from divergences between the target population and the frame population. The target population for the study has been defined as establishments with 5 or more employees in 19 of the 21 NACE Rev.2 sections (all except for NACE T and U).

Theoretically, divergences between the target and the frame population can exist for each single aspect of this definition. The divergences can take various forms, such as:

- An over-coverage in the sampling frame, e.g. due to multiple listings
- Misclassifications
- An under-coverage in the sampling frame.

For ESENER-3, partly address registers of commercial address providers and partly official registers compiled by the national statistical offices were used (usually the registers collected as SBR = Structural Business Register by Eurostat). Official registers are not necessarily preferable over commercial sources. In some countries, they may have a higher overall coverage of the universe than commercial sources, but they also frequently have weaknesses in specific sectors, particularly in NACE K (Finance and insurance), O (Public Administration), P (Education) and Q (Human health and social work activities).

Official company registers are partly also less frequently updated than commercial registers. For some of the official registers telephone numbers are not available and have to be added from other sources. In this step, it is usually not possible to find matches for all addresses. The unmatched addresses without telephone number do not have a chance to be selected for the interview in a telephone survey such as ESENER-3.

Commercial sampling frames, on the other hand, are more likely to have coverage issues in particular sections of the universe (e.g. specific sizes, sectors or ownership types). Some of the commercial frames do also not work with a very clearly defined units and may include a company or establishment several times (e.g. if listed with multiple departments or with slightly different spellings in the sources the frames use for the compilation of their addresses).

What is ultimately the best choice for a country depends on a number of criteria such as the availability of telephone numbers for the addresses, the representativeness of the entries or the accessibility of the registers (in some countries, the official company or establishment registers are not made accessible for survey purposes). The majority of sampling frames used for ESENER-3 was of commercial nature (see Technical Report, Chapter 6.2 for an overview over the registers used)⁶.

Over-coverage

In ESENER-3, an over-coverage in the form of multiple listings of establishments was generally not an issue. If a register has problems with over-coverage, i.e. if it lists the same units more than once without indicating that the addresses belong together, this would become evident in the contact phase of the interview since in most establishments there is only one person that “knows best about health and safety in the establishment”. In most contact attempts at the same address, the interview would thus end up

⁶ In some countries, the distinction between commercial and official registers is however difficult since statistical offices have sometimes outsourced the provision of the business addresses they are collecting to a commercial institute.

with the same person. This person would usually remember the previous call, particularly since it is a rather long interview and not easily to be confused with any commercial survey among businesses. Exceptions to this may occasionally occur where the initial contacts in the organisation (usually the switchboard personnel) identified different target persons for the interview. These cases can however be considered as extremely rare.

Double interviewing due to multiple listings of a unit in the sampling frame is therefore not an issue of concern. Actually, five interview doublets were identified in the ESENER-3 data-checks. But these were all cases where an establishment completed both the telephone and the online interview⁷.

For double calls due to multiple listings of an address, the response code 17 "Already questioned (double address)" was foreseen. The analysis of the gross data shows that this response code was ticked only in a small share of 0,4% of all addresses (n = 1.950 cases) touched for the survey. There are only 7 countries where this share amounts to 1,0% or more: Luxembourg (4,8%), Malta (2,1%), Cyprus (1,7%), Iceland (1,4%), Denmark (1,2%), Belgium (1,2%) and North Macedonia (1,0%). Luxembourg, the country with the by far highest share of double addresses according to this criterion, uses an establishment-level sampling frame based on the national Yellow Pages telephone register. In Yellow Pages registers, double listings tend to occur more frequently, particularly for organisations with different business areas.

Not all addresses coded as "Already questioned (double address)" were actually listed twice or more often in the sampling frame. The code was also used for cases where subsidiaries of multi-site organisations contacted for ESENER re-directed the interview request to their headquarters. If the headquarter itself was also selected and contacted for ESENER or if other subsidiaries had done the same before, it may have declined the further interview request with the argument that it had already been called.

Misclassifications

Misclassifications caused by incorrect information about frame units are an issue that occurs more frequently than over-coverage in registers with addresses of establishments or enterprises. Misclassifications can occur in each element of the definition of the target population:

a) Misclassifications due to a wrong size attribution

The number of workers an organisation employs is subject to changes over time. Even if an address register is updated regularly, the number of employees as listed in the address register may have changed since the last update. Therefore, one of the first questions of the survey (Q102) asked for the number of employees. If a unit turned out to have less than 5 employees at the time of being interviewed, the interview was terminated at that point.

The number of employees as provided in Q102 was taken as base for the size-classification of all interviews in the data-set, replacing the initial information from the address register. The size indication from Q102 was used for all further steps, from the filtering of questions to the monitoring of the sample distribution and the weighting of the data.

Due to this "real time reaction" on wrong size classifications, size misclassifications are not a major issue for the data collected. The data is correctly classified on the number of employees the respondent considered as correct. Even if in some cases in fact the respondent's answer rather than the size

⁷ If an establishment had not yet reacted to the online invitation one week after the second online reminder, it was called again by telephone. If the establishment had completed the online interview between the drawing of addresses for the CATI reminder and the (successful) telephone call, it may have given an additional interview by phone, going on the assumption that the completed online questionnaire got lost due to technical problems. In these cases, the CATI interview was kept while the CAWI interview was deleted from the data-set.

indication from the address source might be wrong, this is not a problematic issue because the size of the unit as indicated by the respondent corresponds to the unit the respondent had in mind when answering the questionnaire.

The overall number of such misclassifications is nevertheless surprisingly high: In total, 24.220 addresses were screened out due to a wrong size indication in the sampling frame and 13.075 establishment turned out to belong to another size-class than the one indicated in the sampling frame⁸ (see chapter 6.1.6 of the Technical Report for more details).

In countries that need to use company addresses, such deviations between the indicated size-class and the size information provided in the interview are to be expected for the multi-site organisations. While the address register here lists the number of employees working in the entire company (including its subsidiaries), the interview is conducted in just one of its local units and the question on the number of employees in the establishment (Q102) is explicitly meant to refer to the selected unit only. This unit may well be in a smaller size-class than the entire company or organisations with all its local units in the country.

For countries using establishment-level sampling frames, the unit listed in the frame and the unit questioned for the survey is however normally the same, even in case of units from multi-site organisations. Nevertheless, for between 24,2% (AT) and 41,8% (UK) of the interviews, the size-class according to the data is different from the size-class indicated in the sampling frame. For some non-screening countries, the reason for these frequent misclassifications is that the information on the size of the subsidiaries is only estimated on base of the total size of the company/organisations and its number of local units⁹.

In a country where many addresses drop out of the universe because they turn out to have less than 5 employees, there is reason to assume that among the addresses with less than 5 employees excluded ex ante for the survey there would also be a notable number of establishments that in fact have 5 or more employees at the time of being interviewed, taking as base the definition of employees from the questionnaire (Q102). For an optimal coverage of establishments just slightly surpassing the 5 employee threshold, it could thus be considered to also sample establishments with less than 5 employees according to the information from the sampling frame for countries where this type of inaccuracies occurs frequently¹⁰.

b) Misclassifications due to a wrong sector attribution

Correct sector attributions in the data-set are as crucial for the data quality of ESENER as correct size attributions. Therefore, at the beginning of the questionnaire (Q108) respondents were asked to confirm whether the sector attribution derived from the information in the sampling frame is correct. If it was considered wrong, the respondent was asked to describe the main activity of the establishment in one keyword (Q109). The interviewer (or the respondent in case of an online interview) then entered the keyword into a search mask and the search engine showed all NACE descriptions that include this keyword, hereby using a combination of the NACE 1-digit (sections) and 3-digit levels (groups) of classification. The interviewer then read out all descriptions shown on the screen, asking the respondent which of these (if any) meets best the activity at the establishment.

⁸ Among the 43.852 valid interviews in the net sample that had a size classification from the address source, 13.075 finally ended up in another size-class when considering the size indication from the respondent (Q102). Size jumping can be observed in both directions: 7.298 addresses had a larger size-class in the address than in the data while for 5.640 addresses the size indicated in the data was larger. In relative terms, in total 28,5% of the interviews (with a size indication from the address) switched the size-class while for 71,5% no such change could be observed.

⁹ Examples for this are the sampling frames used in DE and ES. The methods of estimation hereby vary – some providers just divide the total number of employees in the organisation by the number of its units, others attribute a larger size to the headquarters than to the subsidiaries.

¹⁰ This measure would be costly as a lot of the addresses with a size indication of 1-4 employees in the sampling frame would not be usable because they actually have less than 4 employees also in reality (at the time of the interview). Besides, it would be difficult to steer the sample correctly then because the correct universe is usually not known as for most countries statistics indicating the share of establishments with 5, 6, 7, 8 or 9 employees are not available.

If no appropriate descriptions could be found by that way, the respondent was subsequently asked (Q110) to describe the main activity in a few words so that these could be attributed afterwards to the appropriate NACE category. After finalisation of the survey, these verbatims were translated into German and were centrally coded by a specialized coding team at Kantar Public Division.

All in all, 9,5% of respondents had considered the sector indication from the address source as wrong, with national results ranging from 2,6% wrong codes in Estonia to 14,8% in Austria. This is considerably less than the 15,6% observed in ESENER-2¹¹.

For about a quarter (26%) of the interviews where the sector indication from the frame was considered wrong, the coding of the sector description provided by the respondent did however lead to exactly the same NACE Rev.2 2-digit code that was indicated in the address source. Furthermore, some of the codes that needed corrections on the NACE Rev.2 2-digit level turned out to be still within the same NACE Rev.2 section indicated in the address source. For the steering of fieldwork and for analysis, this is the finest level of sector differentiation usually used for ESENER. Considering this, in the end only 5,8% of all sector codes had to be revised on base of the verbatim sector descriptions provided by the respondents. More details on the sector coding and its results can be found in the Technical Report, chapters 6.1 and 11.3.

Summing up, it can be said that the procedure set up in the questionnaire for the verification and correction of the sector of activity and the later sector coding ensure that the sector attributions available in the data-set are highly reliable. The switches between different NACE Rev.2 1-digit sectors led to some additional discrepancies between the targeted and the finally achieved sector structures, but only for the relatively few interviews where a manual subsequent coding was necessary.

c) Misclassifications due to a wrong unit classification (enterprise vs. establishment)

A third type of misclassifications is owed to possibly erroneous classifications of some addresses as “establishments” or “local units”. Following the recommendations of EUROSTAT, the term “establishment” was defined for the survey as “...*(a)n enterprise or part of an enterprise that is situated in a single location and in which only a single productive activity is carried out or in which the principal productive activity accounts for most of the value added*”¹².

Not all establishment registers use the same logics and definitions for distinguishing “local units” or “establishments” from larger units such as “companies” or “enterprises” or from smaller units such as “departments”. Larger enterprises have for example often split their different businesses into various legally independent units that are located at the same geographical site, are often even doing basically the same things and are thus not easily recognizable as firms of their own. These may in some registers be listed as local units, in others as companies (because of their legal independence) and in still others they might not be separately listed at all but considered as part of the organisation at that place.

This type of misclassification of a unit cannot easily be controlled for in a sample survey and it is likely to occur in part of the interviews.

Under-coverage

Under-coverage is certainly the most important frame error that may occur in a sample survey among organisations. It results from incomplete address sources.

¹¹ These improvements are largely owed to changes in the sampling frames used for the survey. The commercial frame used for CH in ESENER-2, for example, had the highest rate of sector switchers in ESENER -2, with almost a quarter (24,2%) of addresses being wrongly classified. In ESENER-3, the official BUR register was used as sampling frame and the share of establishments with a wrongly coded sector dropped to 6,9%.

¹² See <https://ec.europa.eu/eurostat/statistics-explained/index.php/Glossary:Establishment>

For the survey, the best available establishment-level sampling frames suitable for a survey such as ESENER were selected. Most of these are address registers compiled by private organisations because most national statistical offices do not maintain address registers at the establishment level or they do not make them accessible for private survey purposes. Several address registers used for the survey are therefore not based on obligatory entries of all organisations, but on data compiled from a number of publicly accessible sources such as trade registers, tax registers etc., sometimes amended by own investigations or address verifications on part of the frame provider. In these sources, some types of organisations are usually listed more completely than others. While e.g. large commercial organisations tend to be listed well, entries of smaller units and entries of public or non-profit units are often less complete.

Where this type of under-coverage stemming from incomplete entries in the address registers is constrained to the under-coverage of specific size classes or sectors of activity, it is largely compensated for by the weighting procedure and thus not a problem for the survey quality.

Under-coverage due to criteria not clearly linked to either size or sector may be a source for major inaccuracies in the survey measurement. This is, for example, the case for the non-coverage or under-coverage of public institutions within a sector characterized by a mixture of public, private and non-profit institutions. Another example would be the under-coverage of family-owned businesses among the smaller units, occurring particularly in commercial sampling frames compiled from published company information.

This type of under-coverage is hard to identify. In most countries, no reference statistics are available that show the real distribution of the universe by establishment size, sector and ownership type. The share between public and private ownership within a sector, for example, may vary largely between countries, depending on how services such as public transport or the health and social work system are organised. In view of the lack of reliable and sufficiently detailed statistical data on the ownership type, it is not possible to correct such disproportionalities ex post by way of an additional weighting factor.

In the ESENER-3 data-set, there are some larger differences regarding the ownership type (public vs. private) in NACE P (Education) and Q (Health and social work activities):

- In Malta and Romania, only a small share of the interviews in NACE P was made with public entities (public schools etc.). There is probably some undercoverage of public entities in this sector for these two countries.
- In Cyprus, Portugal and Finland, only a small share of interviews in the health and social work sector was done with public units. But whereas probably in all European countries the majority of establishments in NACE P (Education) are in public ownership, national health systems are likely to differ much more in this aspect¹³.

Whereas a certain undercoverage of public units is evident in some countries at least for NACE P, it is not always straightforward to identify the reasons for the undercoverage. The main reason for an undercoverage in the net sample is not necessarily a problematic sampling frame. In some cases, orders from the “headquarters” of public entities with a large number of local units not to participate in any surveys may also lead to a low response rate for public entities¹⁴.

¹³ For Germany, for example, the rather low share of public units in NACE Q is plausible since in Germany, the health system is based on a large number and variety of mostly relatively small private medical surgeries. In Finland, in turn, there is a strong public health system, but this is based on relatively few (mostly quite large) health centres, in total about 270 in the entire country. In addition to these, there are also private surgeries. Though these are overall much less important for the health provision to the citizen, the number of these surgeries is much larger than the number of publicly owned health centres (for the Finnish health system see: https://www.ewi-psy.fu-berlin.de/einrichtungen/arbeitsbereiche/ppg/aservice/newsletter/IPG-newsletter_archiv/IPG-NL-02-04/Gesundheitssystem_Finnland/index.html). This example shows that it is very difficult to assess a sampling frame or a net sample in a cross-national survey such as ESENER on criteria such as the ownership type.

¹⁴ For ESENER-3, for example, the fieldwork partner in Greece informed about the necessity to get an official recommendation from the Ministry of Education for doing interviews in public schools since the management of public schools is advised not to participate in any surveys. EU-OSHA and Kantar applied for such a recommendation, though with limited success.

Table 2: Establishments in the education and the health sector, by ownership type

Country	Public sector (Q111)?							
	NACE P (Education)				NACE Q (Health and social work activities)			
	Yes	No	No answer	Total	Yes	No	No answer	Total
at	57,3%	41,7%	1,0%	100,0%	34,6%	65,4%	0,0%	100,0%
be	71,4%	27,8%	0,8%	100,0%	27,0%	73,0%	0,0%	100,0%
bg	82,6%	17,4%	0,0%	100,0%	59,3%	40,7%	0,0%	100,0%
ch	67,6%	32,4%	0,0%	100,0%	38,1%	61,3%	0,6%	100,0%
cy	52,5%	45,8%	1,7%	100,0%	14,3%	85,7%	0,0%	100,0%
cz	91,1%	8,4%	0,5%	100,0%	61,9%	38,1%	0,0%	100,0%
de	67,9%	31,4%	0,6%	100,0%	25,4%	73,5%	1,1%	100,0%
dk	81,4%	18,6%	0,0%	100,0%	82,3%	17,7%	0,0%	100,0%
ee	83,1%	16,9%	0,0%	100,0%	54,8%	45,2%	0,0%	100,0%
el	75,7%	24,3%	0,0%	100,0%	36,3%	63,8%	0,0%	100,0%
es	47,6%	52,4%	0,0%	100,0%	20,3%	78,5%	1,2%	100,0%
fi	64,8%	35,2%	0,0%	100,0%	19,1%	80,9%	0,0%	100,0%
fr	79,1%	20,9%	0,0%	100,0%	30,6%	69,4%	0,0%	100,0%
hr	62,7%	35,6%	1,7%	100,0%	46,8%	53,2%	0,0%	100,0%
hu	36,4%	63,6%	0,0%	100,0%	37,4%	61,6%	1,0%	100,0%
ie	44,1%	54,4%	1,5%	100,0%	27,8%	70,7%	1,5%	100,0%
is	79,5%	20,5%	0,0%	100,0%	55,4%	41,9%	2,7%	100,0%
it	31,9%	68,1%	0,0%	100,0%	43,8%	56,2%	0,0%	100,0%
lt	94,4%	5,6%	0,0%	100,0%	82,4%	17,6%	0,0%	100,0%
lu	61,3%	38,7%	0,0%	100,0%	30,3%	69,7%	0,0%	100,0%
lv	80,3%	19,7%	0,0%	100,0%	70,3%	29,7%	0,0%	100,0%
mk	70,3%	29,7%	0,0%	100,0%	58,5%	41,5%	0,0%	100,0%
mt	10,7%	89,3%	0,0%	100,0%	30,4%	69,6%	0,0%	100,0%
nl	72,2%	25,0%	2,8%	100,0%	24,8%	74,3%	0,9%	100,0%
no	71,7%	28,3%	0,0%	100,0%	66,3%	33,4%	0,3%	100,0%
pl	91,1%	8,9%	0,0%	100,0%	73,0%	25,2%	1,7%	100,0%
pt	61,9%	38,1%	0,0%	100,0%	14,3%	85,7%	0,0%	100,0%
ro	31,4%	68,6%	0,0%	100,0%	25,8%	74,2%	0,0%	100,0%
rs	77,2%	22,8%	0,0%	100,0%	48,6%	51,4%	0,0%	100,0%
se	81,8%	18,2%	0,0%	100,0%	72,4%	27,6%	0,0%	100,0%
si	93,8%	6,2%	0,0%	100,0%	77,4%	22,6%	0,0%	100,0%
sk	88,6%	10,0%	1,4%	100,0%	63,0%	34,8%	2,2%	100,0%
uk	52,5%	45,7%	1,9%	100,0%	26,1%	71,6%	2,3%	100,0%
ALL	69,0%	30,6%	0,4%	100,0%	39,5%	59,8%	0,7%	100,0%

4.1.2 Non-response errors I: Unit non-response

The achievement of high cooperation and response rates is an important goal for any survey with high quality requirements. Though the avoidance of any response biases is more important than the achievement of the highest possible response rate, with higher overall response rates the risk of any response biases can in tendency also supposed to be lower.

The response and cooperation rates show the share of completed interviews, though with different bases:

- The base for the calculation of the response rate is the sum of all potentially eligible addresses used for the survey. This includes addresses that could not be reached during fieldwork time (Wrong telephone number, No answer, Answering device etc.) but it usually excludes units that turned out to be out of scope of the survey (private households, establishments with < 5

employees, extinguished businesses etc.). The treatment of units for which it is unclear whether or not they are within the scope of the survey (unknown eligibility) differs between variants of response rates.

- The base for the calculation of the cooperation rate is the sum of all addresses that could be successfully contacted for the survey. Again, units clearly out of scope are excluded and for those with an unknown eligibility, different variants of cooperation rates exist.

In a multi-country survey like ESENER-3 that has to use sampling frames with different quality levels, an outcome rate taking into account the share of wrong addresses or ineligible units in the calculation is better suited for the assessment of both the success of fieldwork itself and the probability of having any systematic (non-)response biases.

Non-responses due to wrong telephone numbers, fax numbers (instead of telephone numbers) etc. as well as non-participation due to ineligibility are mostly a sign of the quality of the address registers available for the survey and not primarily of fieldwork quality. In Response rate 3 (see Table 3), the ineligible cases and a share of addresses with an alleged ineligibility are excluded. In Cooperation rate 3 modified, additionally also all addresses not reached during the fieldwork period are excluded. The latter thus marks the most "benevolent" of the range of outcome rates.

The table below (Table 3) illustrates the mode of calculation for the cooperation and the response rate in different variants.

Table 3: Definition of response, cooperation and other rates

Code No.	Response code description (hierarchical code)	AAPOR Response Rate 1	AAPOR Response Rate 3	AAPOR Cooperation Rate 1	AAPOR Cooperation Rate 3 modified	AAPOR Cooperation Rate 3
1	No answer			not eligible	not eligible	not eligible
2	Answer device			not eligible	not eligible	not eligible
3	Busy			not eligible	not eligible	not eligible
4	Information tone - Fax - Modem			not eligible	not eligible	not eligible
5	Wrong telephone number			not eligible	not eligible	not eligible
6	Callback possible (general appointment)					not included
7	Definitive appointment with target person					not included
8	Refusal by target person					
9	Refusal by contact person/reception (upfront refusal)					
13	No establishment at this address (private household etc.)	not eligible	not eligible	not eligible	not eligible	not eligible
14	Inactive establishment, terminated	not eligible	not eligible	not eligible	not eligible	not eligible
17	Already questioned (double address)	not eligible	not eligible	not eligible	not eligible	not eligible
18	Complete CATI interview					
21	Stratification maximum reached (cell full)			not eligible	not eligible	not eligible
22	Complete, but unanalyzable (data/interviewer issues detected during fieldwork or in final data cleaning)					
34	Refusal - add number to DO NOT CALL LIST					
35	Partial interview, to be called back					
36	Partial interview, not to call back					
37	No appointment with target person possible during fieldwork time and period			not eligible	not eligible	not eligible
38	Target person does not speak proposed languages			not eligible	not eligible	not eligible
42	Size out of target (less than 5 employees or NA in Q105)	not eligible	not eligible	not eligible	not eligible	not eligible
43	Refusal to give information in Q111, Q112 (sector information)					
44	No single establishment with 5 or more employees (Q051=0)	not eligible	not eligible	not eligible	not eligible	not eligible
45	Size of first contact out of scope but interview possible at subsidiary (screening countries)	not eligible	not eligible	not eligible	not eligible	not eligible
46	Interview terminated after screening phase, not to call back					
51	Online invitation with no result					not eligible
52	Complete CAWI interview					
53	Online interview rejected due to quality reasons					
56	No adequate target person at the establishment			not eligible	not eligible	not eligible
58	Online refusal					

= complete interview (denominator)
 = included in denominator, unknown eligibility
 = included in denominator
 = excluded

The country differences have various reasons. The general attitude of potential respondents towards surveys conducted at the workplace is an important factor, influenced among others by the number of survey requests an average organisation is facing per year. But other factors, such as the importance of the topic in national debates or the experience and quality of the interviewer teams, also play a role.

Cooperation and response rates in ESENER-3 vary also considerably between size-classes and sectors. Over all countries, it was generally more problematic to reach the very small establishments than the large ones. But part of the difficulties with small establishments were frequent misclassification by size, causing the dropout of many of the successfully contacted units as not pertaining to the survey universe.

In terms of sectors of activity, the Public Administration (NACE O) stands out with the highest cooperation rate and one of the highest response rates. NACE I (Accommodation and food service activities), in turn, has the lowest response rate and is in all calculation variants also among the countries with a comparatively low cooperation rate (see Chapter 7.5 of the Technical Report for more details about cooperation and response rates).

Table 4: Cooperation and response rates by country, different calculation variants

Country	Eligibility rate	RR1	RR3	CR1	CR3 modified	CR3
AT	67%	10%	14%	13%	19%	16%
BE	67%	18%	25%	27%	36%	39%
BG	54%	11%	19%	18%	28%	21%
CH	63%	29%	39%	39%	50%	59%
CY	31%	15%	36%	22%	47%	27%
CZ	46%	3%	7%	5%	10%	7%
DE	25%	5%	18%	6%	21%	9%
DK	72%	18%	24%	28%	35%	45%
EE	70%	36%	45%	48%	57%	60%
EL	69%	15%	20%	19%	26%	25%
ES	44%	6%	14%	12%	23%	15%
FI	66%	24%	33%	33%	42%	46%
FR	73%	15%	19%	21%	26%	34%
HR	70%	11%	15%	14%	19%	18%
HU	67%	10%	14%	19%	26%	25%
IE	85%	23%	25%	27%	30%	57%
IS	70%	32%	40%	42%	51%	54%
IT	80%	14%	17%	18%	22%	24%
LT	73%	23%	29%	31%	38%	50%
LU	51%	20%	34%	28%	43%	39%
LV	55%	26%	38%	30%	44%	48%
MK	80%	55%	60%	69%	73%	85%
MT	43%	17%	33%	27%	45%	40%
NL	50%	10%	19%	14%	24%	18%
NO	58%	11%	17%	18%	27%	26%
PL	43%	9%	18%	12%	25%	14%
PT	66%	18%	25%	26%	35%	40%
RO	61%	8%	13%	15%	23%	19%
RS	59%	26%	38%	48%	61%	66%
SE	71%	19%	26%	24%	31%	43%
SI	88%	24%	26%	30%	33%	39%
SK	46%	3%	7%	5%	10%	6%
UK	63%	19%	27%	24%	33%	41%
ALL	56%	11%	18%	16%	25%	22%

Possible response biases due to selective non-response by sub-groups of the universe

The cooperation and response rates achieved for a survey give an indication of the amount of unit nonresponse. But though high cooperation and response rates are generally desirable, they are not the only and even not the most important indication of non-response errors. Even more important than achieving high cooperation and response rates is the avoidance of any response bias, i.e. of selective unit nonresponse. Selective unit nonresponse occurs if specific sub-groups of the target population systematically participate to a lower degree than other sub-groups. Selective non-response and thus potential non-response biases can have various causes. As for ESENER-3, the most likely causes for selective non-response are:

- a. Non-response related to difficulties in reaching sub-groups of the frame population
- b. Non-response related to the subject of the survey

a) Non-response related to difficulties in reaching sub-groups of the frame population

Within the sampling and weighting design chosen for ESENER-3, selective non-response related to a lower participation of establishments of particular sizes and sectors are automatically corrected as long as the selectivity is confined to the size and sector groups defined by the sampling matrix (4 size-classes and 8 sector groups): If addresses for a particular cell were exhausted, then new addresses were provided just for this segment. An establishment that refused participation was thereby in principle replaced by another establishment of the same sector and size-class and the originally envisaged number of interviews could be reached in spite of the selective non-response. Where such replacements with new addresses were necessary, these were drawn on base of a 76-cell matrix, defined by the 4 size-classes used in the survey and the 19 relevant NACE sections.

Moreover, for the weighting also this finer differentiation by 76 was applied in all countries. In the weighing process the achieved number of interviews in a cell was basically put into ratio with the real number of establishments in that cell (universe)¹⁵. By this way, unequal non-responses within the different cells of the sampling and weighting matrix are equaled out. A pre-condition for this mechanism is however that - in spite of the higher non-response – there are still enough interviews available for the concerned sectors.

For unequal non-response within the cells of the weighting matrix, the situation is different. For this type of non-response – e.g. unequal response behaviour of public and private institutions within the Health and social services sector – a correction is hardly possible because in most countries reliable statistics on the distribution of units according to this criterion are not available.

Feedback of the fieldwork institutes suggests for example that particularly within the retail trade sector, subsidiaries were often generally not authorized by their headquarters to take part in any survey and were told to either refuse or to direct the interview to the headquarters. Likewise, small subsidiary units of the Public Administration (e.g. local police stations) in some countries tended to refuse participation because they felt not entitled to answer the survey. A consequence of this type of selective non-response is that within these sectors and size-classes, single-site units and headquarters tend to be over-represented while the situation in the subsidiaries is less reflected by the survey results. The magnitude of this type of non-response cannot precisely be quantified because most establishment-level address sources do not provide any information on whether an address belongs to the headquarters or to one of the subsidiaries.

Another example are private doctor's surgeries in some countries. These were partly difficult to get for the survey because these surgeries are frequently approached by survey requests for the pharmaceutical industry for which they are used to get high incentives.

b) Non-response related to the subject of the survey

Non-response by certain groups of the investigated population that are related to the subject of the survey are likely to occur to a certain degree in a survey like ESENER-3, asking organisations about issues that are "good practice" for a good employer, if not even legal obligations. In the interviewer instructions and in the formulation of the survey entry care has been taken to encourage all types of employers to participate, those performing well in health and safety matters as well as those that are aware of having certain deficits in this area.

The reassurance of strict confidentiality and anonymous data analysis, supported by an explicit data protection sheet sent to all interested respondents also contributes to avoiding a bias in this regard, with those performing well in OSH participating more than others. In how far these efforts were successful can only be assessed by researchers and experts on the local level, e.g. by comparing certain findings

¹⁵ This is a very simplified description of the weighting. For a detailed description of the applied GREG weighting method see chapter 9 of the Technical Report.

(e.g. on the share of establishments carrying out risk assessments) with possibly available statistics of the labour inspectorate or other national sources. Some comparisons for validation can also be made by comparing the data (in the employee proportional weighting) with data of Eurofound's Working Condition Survey, e.g. on OSH training provided to employees.

In the ESENER-3 pilot survey, efforts were made to learn more about whether and to which extent such a self-selection bias occurs. To this end, a set of three statements trying to measure the OSH culture in the establishment was not only asked to respondents of the full survey, but also to the contact persons (gatekeepers or targeted respondents) in establishments refusing to participate in the full survey. A comparison between the answers of both groups has shown that in tendency, those having refused participation in the full survey indicated to have an even slightly better OSH culture than those participating in the full survey. This finding is however not very robust: On the one hand, only a small share of the non-respondents was willing to answer the question. On the other hand, the measurement quality of the three statements was problematic, with respondents interpreting the statements in different ways. The questions were thus dropped for the main survey.

4.1.3 Non-response errors II: Item non-response

The share of unspecific answers to questions (Don't know/No answer) is often cited as a measure for the data quality. Indeed, a high share of item non-response for a question indicates that it was frequently not well understood or that the respondent understood the question but was not able or willing to provide an answer.

The share of item non-response in ESENER-3 is generally low: Calculated per interview, the item non-response rate in the data-set is 1,5%, national boost questions not taken into account. This means that on average, a respondent did not answer 1,5% of the questions¹⁶ he or she received.

Table 5 below shows that for the single questions, item non-response rates are ranging from 0,0% in the open-ended sector question Q110 to 21,9% in Q109, the question asking to describe the sector of activity in one keyword for the sector search engine included in the CATI/CAWI script. Both questions are however atypical for the survey: Q110 is the only open-ended question. For Q109, the questions with the sector search engine, the item non-response rate does not refer to the answers of the respondents. It reflects the share of respondents for which there were either no or no adequate sector descriptions provided by the search engine after having entered the sector keyword. In these cases, interviewers were advised to tick "Don't know/No answer" in order to get to the follow-up question Q110 asking these respondents to describe the sector in some more words for a posterior manual sector coding.

For the remaining questions, item non-response rates vary from 0,1% in Q310_2 to 6,8% in Q159 (development of the sickness rate) and Q260_3 (risk assessments considered as too burdensome). Further questions with a high item non-response rate above 5% are Q307 (assessment on whether psycho-social risks are more difficult to tackle than other risks, 5,8%) and Q358_1 (employers organisations as source for health and safety information used at the workplace, 5,6%).

For Q112, the question asking for the year of foundation of the establishment, an elevated item non-response had been anticipated and people answering with "Don't know" in Q112 were additionally asked Q112x which repeated the question but offered categories instead of asking for an open numerical answer. After offering categories, the initial non-response rate of Q112 dropped to 0,6% if taking the answer of both Q112 and Q112x together.

For the CAWI interviews, the item non-response rate is higher than for CATI: Taking all CAWI interviews into consideration (i.e. including the 376 CAWI interviews not accepted for the data-set), the item non-response share for CAWI is 6,0%. This value is however influenced by the high item non-response value

¹⁶ In questions with item batteries where each item had to be answered separately, each item was counted as one question for this analysis.

of 23,5% in the rejected interviews. Among the 2.166 accepted CAWI interviews, the average item non-response is at 3,0%. The item non-response rate for the CATI interviews is just 1,4%. For CAWI, the rate of 3,0% is however still very low, suggesting that the questionnaire was all in all easy to answer.

Table 5: Share of item non-response per question (unweighted results)

Question	No. of answers (unweighted)	No. of item non- response (NA/DK*)	Share of non- response (NA/DK) in %	Question	No. of answers (unweighted)	No. of item non- response (NA/DK*)	Share of non- response (NA/DK) in %
Q100	28.818	68	0,2%	Q202_1	45.420	240	0,5%
Q100all	45.420	105	0,2%	Q202_2	45.420	445	1,0%
Q101a	11.470	197	1,7%	Q202_3	45.420	645	1,4%
Q101b	2.390	34	1,4%	Q202_4	45.420	332	0,7%
Q101all	13.860	231	1,7%	Q202_5	45.420	1.803	4,0%
Q102gr	45.420	-	0,0%	Q250	45.420	706	1,6%
Q103	45.420	198	0,4%	Q251	35.782	167	0,5%
Q104	45.420	98	0,2%	Q252_1	35.782	619	1,7%
Q105	45.420	1.146	2,5%	Q252_2	15.672	143	0,9%
Q106	45.420	153	0,3%	Q252_3	35.782	600	1,7%
Q107	45.420	261	0,6%	Q252_4	35.782	636	1,8%
Q108	45.420	139	0,3%	Q252_5	35.782	1.094	3,1%
Q109	4.302	940	21,9%	Q252_6	35.782	722	2,0%
Q110	940	-	0,0%	Q252_7bo	391	7	1,8%
Q111	45.420	236	0,5%	Q252_8bo	735	-	0,0%
Q112x	2.934	141	4,8%	Q252_9bo	1.208	15	1,2%
Q112gr	35.705	209	0,6%	Q253	5.483	176	3,2%
Q113	45.420	451	1,0%	Q254	16.491	283	1,7%
Q114	22.146	348	1,6%	Q255	14.438	189	1,3%
Q150	45.420	207	0,5%	Q256	35.782	1.492	4,2%
Q151_1	45.420	502	1,1%	Q256gr	35.782	1.492	4,2%
Q151_2	45.420	974	2,1%	Q257	35.647	869	2,4%
Q151_3	45.420	878	1,9%	Q258	35.782	533	1,5%
Q151_4	45.420	871	1,9%	Q259bo	1.663	-	0,0%
Q151_5	45.420	1.045	2,3%	Q260_1	8.932	267	3,0%
Q152	45.420	840	1,8%	Q260_2	8.932	77	0,9%
Q153	29.251	331	1,1%	Q260_3	8.932	603	6,8%
Q154	45.420	2.258	5,0%	Q260_4	8.932	345	3,9%
Q155	45.420	688	1,5%	Q261	8.932	200	2,2%
Q156	40.967	162	0,4%	Q262_1	45.420	507	1,1%
Q157_1	45.420	355	0,8%	Q262_2	45.420	598	1,3%
Q157_2	45.420	291	0,6%	Q262_3	45.420	885	1,9%
Q157_3	45.420	195	0,4%	Q262_4	45.420	354	0,8%
Q157_4	45.420	261	0,6%	Q262_5	45.420	513	1,1%
Q158	45.420	268	0,6%	Q263_1	45.420	346	0,8%
Q159	45.420	3.080	6,8%	Q263_2	45.420	539	1,2%
Q160	45.420	1.234	2,7%	Q263_3	45.420	353	0,8%
Q161	11.879	444	3,7%	Q263_4	45.420	384	0,8%
Q162	20.870	226	1,1%	Q263_5	45.420	477	1,1%
Q163	20.870	493	2,4%	Q263_6	45.420	396	0,9%
Q164a	26.700	158	0,6%	Q263_7	45.420	759	1,7%
Q164b	18.563	58	0,3%	Q300	20.870	727	3,5%
Q164all	45.263	216	0,5%	Q301	20.870	465	2,2%
Q200_1	45.420	122	0,3%	Q302	13.526	351	2,6%
Q200_2	45.420	238	0,5%	Q303a	20.870	841	4,0%
Q200_3	45.420	102	0,2%	Q303b	24.550	511	2,1%
Q200_4	45.420	456	1,0%	Q303all	45.420	1.352	3,0%
Q200_5	45.420	91	0,2%	Q304_1	45.420	1.145	2,5%
Q200_6	45.420	164	0,4%	Q304_2	45.420	1.092	2,4%
Q200_7	45.420	91	0,2%	Q304_3	45.420	886	2,0%
Q200_8	45.420	116	0,3%	Q304_4	45.420	1.699	3,7%
Q200_9	45.420	95	0,2%	Q304_5	45.420	1.108	2,4%
Q200_10	45.420	126	0,3%	Q305	38.925	605	1,6%
Q201_1	45.420	246	0,5%	Q306	38.925	1.594	4,1%
Q201_2	45.420	328	0,7%	Q307	36.154	2.108	5,8%
Q201_3	45.420	660	1,5%	Q308_1	9.023	199	2,2%
Q201_4	45.420	250	0,6%	Q308_2	9.023	153	1,7%
Q201_5	45.420	118	0,3%	Q308_3	9.023	148	1,6%
Q201_6bo	3.018	7	0,2%	Q308_4	9.023	152	1,7%
Q201_7bo	3.018	14	0,5%	Q309	35.782	2.354	6,6%
Q201_8bo	3.018	19	0,6%	Q310_1	45.420	78	0,2%

Question	No. of answers (unweighted)	No. of item non- response (NA/DK*)	Share of non- response (NA/DK) in %	Question	No. of answers (unweighted)	No. of item non- response (NA/DK*)	Share of non- response (NA/DK) in %
Q310_2	45.420	43	0,1%	Q355_4	25.164	214	0,9%
Q310_3	45.420	111	0,2%	Q355_5	45.420	420	0,9%
Q310_4	45.420	278	0,6%	Q355_6	23.209	853	3,7%
Q310_5	45.420	222	0,5%	Q356	4.511	75	1,7%
Q310_6	45.420	148	0,3%	Q357	45.420	415	0,9%
Q311	43.244	897	2,1%	Q358_1	45.420	2.533	5,6%
Q312_1	11.947	308	2,6%	Q358_2	45.420	1.447	3,2%
Q312_2	11.947	236	2,0%	Q358_3	45.420	898	2,0%
Q312_3	11.947	101	0,8%	Q358_4	45.420	1.574	3,5%
Q312_4	11.947	134	1,1%	Q358_5	45.420	1.341	3,0%
Q312_5	11.947	114	1,0%	Q358_6	45.420	1.770	3,9%
Q312_6	11.947	180	1,5%	Q359bo	3.018	160	5,3%
Q312_7	11.947	258	2,2%	Q360bo	3.018	121	4,0%
Q312_8	11.947	184	1,5%	Q361bo	3.018	103	3,4%
Q350_1	42.399	754	1,8%	Q362bo	1.999	130	6,5%
Q350_2	40.880	483	1,2%	Q363bo	1.999	760	38,0%
Q350_3	43.601	512	1,2%	Q364_1bo	255	32	12,5%
Q350_4	45.420	416	0,9%	Q364_2bo	255	18	7,1%
Q351	27.868	828	3,0%	Q364_3bo	255	16	6,3%
Q352	32.065	365	1,1%	Q364_4bo	255	12	4,7%
Q353	29.057	262	0,9%	Q364_5bo	255	23	9,0%
Q354	27.868	569	2,0%	Q364_6bo	255	17	6,7%
Q355_1	45.420	573	1,3%	Q400	45.420	1.341	3,0%
Q355_2	18.151	220	1,2%	Q401	45.420	1.800	4,0%
Q355_3	45.420	849	1,9%				

4.1.4 Sampling errors

In a probability sample survey, a selection of units is investigated with the aim of drawing conclusions from this selection to the true situation in the whole universe (here: all establishments with 5 or more employees in the 33 participating European countries). The larger the sample, the lower the sampling error is. But the size of the total net sample is not the only criterion influencing the degree of sampling errors. The sample composition is also important, i.e. the degree to which it represents different sub-groups of the total population under investigation. For each sub-group of the universe that is important to distinguish in the analysis, a sufficiently large number of interviews needs to be available in order to avoid conclusions being drawn on a very weak empirical basis.

The sampling method

The sampling method applied for ESENER-3 can be characterized as a multi-stratified random sampling procedure. In this, the universe is being divided into several strata. In case of ESENER-3, these are defined by country, size and sector (4 size classes and 19 sectors, in total 76 strata per country). For each stratum, the targeted number of interviews was defined in the following way:

- For each size-class, the number of interviews to be targeted was calculated in a mixture of establishment and employee proportionality¹⁷.
- Within the size-classes, the sector targets were then set in an establishment proportional way. The only exception to this is Ireland where strong sector disproportionalities were introduced for

¹⁷ The target for each of the four size classes was determined by its percentage share of the total universe of establishments multiplied by the factor 2 plus the percentage share of the total universe of employees. The sum of these 3 percentage values was then divided by 3 (e.g.: if in country X 40% of all establishments within the defined universe have 5 to 9 employees and 20% of all employees within the defined universe are working in establishments with 5 to 9 employees, then the total share for the size-class was calculated as $(40\% \cdot 2 + 20\%) / 3 = 33,3\%$).

the additionally ordered boost samples (all 1.250 boost interviews had to be made in NACE H and I).

Within each stratum, samples were drawn at random from the selected address registers. The ratio of gross addresses per targeted net interview hereby varied between countries, sectors and sizes and depended on the response rate achieved per cell in ESENER-2 in each country. If a local fieldwork team ran out of addresses for particular cells, then new addresses were provided as replacement for these cells.

The main advantage of this sampling method over a simple random procedure without stratification is that it ensures the availability of enough addresses for each segment within the defined matrix. In a standard, unstratified random sampling procedure, the majority of addresses drawn would be from small establishments with less than 50 employees since these represent the vast majority of units within the defined universe. For the analysis of the situation in larger size-classes, the empirical basis would be very weak. In the disproportionally stratified sample, the size distribution is more even.

The drawback of the stratified sampling procedure in terms of sampling accuracy is that the more even distribution of interviews over the size-classes leads to large establishment-proportional weighting factors in the small size-classes since the majority of establishments is rather small. In statistical terms, the degree of insecurity of the extrapolation from the situation in relatively few establishments with 5 to 9 employees to all establishments of that size is thus higher. Particularly in view of the fact that the relatively few large establishments employ a large share of employees, this drawback seems to be tolerable.

Sampling errors resulting from the screening process

In 17 of the 33 countries, address registers listing the single establishments or local units of multi-site enterprises were not available. Therefore, company/enterprise-level addresses had to be used instead. In the interview, these were first screened for the number of local units they consist of and one of these units was randomly chosen for the further interview. This procedure guaranteed that in all countries, interviews were related to the same type of unit (a selection of headquarters and subsidiaries). The screening leads however to the selection of not more than maximum three units (usually the headquarters and two of its subsidiaries) for each multi-site organisation, while in countries using establishment-based registers far more than three units may be selected for interview within an organisation that has numerous sites. Thus, subsidiary units of multi-site enterprises are likely to be under-covered in the screening countries. This is particularly an issue for sectors of activity characterised by a high share of multi-site organisations, such as “Financial and insurance activities” (NACE K) or “Retail trade” (part of NACE G).

This under-coverage was anticipated and measures were taken to correct for it. This was in a first step done by taking the (mostly estimated) figures on the universe of establishments (not companies) as base for the calculation of the targets within each cell. Secondly, the weighting contributed to a better representation of the subsidiaries in screening countries: subsidiaries of multi-site organisations received an additional weight (a so called ‘entry factor’) in order to compensate for these differences in the selection probabilities. The entry weight is calculated on base of the information about the number of subsidiaries an organisation has in total (within a chosen sector). In order to avoid single interviews (with a maybe untypical answer pattern) to influence results too much, the entry weight was however limited to the factor ‘5’ as maximum per interview in a subsidiary. The entry weight is integrated into the different varieties of establishment-proportional weighting factors delivered with the dataset, it is not a separate factor.

Sampling rates

The 33 economies covered by the survey differ largely in size: In Iceland, the universe of establishments 5+ to be covered by the survey consists of just about 6.000 establishments. In Germany, the largest EU

economy, it amounts to 1,2 million establishments and is about 200 times larger than the universe in Iceland.

Overall, the ESENER-3 universe is estimated to consist of almost 6.4 million establishments. Data on the total universe of establishments, including those with just 1 to 4 employees, were not collected systematically so that it is not possible to exactly determine the share of all establishments this universe covers.

The sample size targeted for ESENER-3 was of 45.200 interviews. The targeted sample size was reached in almost all countries¹⁸ and in a number of countries it was even surpassed so that in total, 45.420 validated interviews are available. With this, ESENER-3 can be considered as a large-scale survey. There are not many cross-national European surveys among organisations that have larger samples.

The overall sampling rate in ESENER-3 was 0,007. Or, in another perspective: **On average**, every interview in ESENER-3 represents about 140 establishments.

The national sample sizes were to a certain degree adapted to the size of the respective country and varied between 450 interviews in Malta to 2.250 in the largest economies. For Slovenia, Norway and Ireland, additional sample boosts came on top.

In spite of the adaptation of sample sizes to the sizes of the economies, the sampling rate varies largely between countries. It is smallest in Germany, the largest country in the sample, with a sampling rate of 0,002 and it is highest in Iceland, with 0,126. Expressed again in another perspective, one interview in Germany on average¹⁹ represents 533 establishments while one interview in Iceland represents just 8 establishments. An interview done in Germany thus on the average represents 67 times as many establishments as an interview done in Iceland.

The precision of the survey in statistical terms does therefore also vary between countries. This large variance by countries is mainly owed to two considerations:

- The staggered sample sizes as applied to ESENER lead to different degrees of precision for the national samples. But if adapting sample sizes more to the size of the various national economies, only very small sample sizes would result for the smaller countries. For these countries, then the total number of observations and particularly the number of observations per cell of the sampling matrix would be very low, rendering national analyses problematic. In the sizing of the national samples, therefore always a tradeoff has to be made in a survey including economies of very different sizes.
- The survey is financed from EU funds. Against this background, it would be difficult to justify a considerably larger spread in the sizes of the national samples.

¹⁸ In Croatia, the final net sample remained 10 interviews below the target, in Portugal 7 and in Ireland 1. These small deviations were agreed with EU-OSHA in an effort to finalise fieldwork on time.

¹⁹ In reality, the rate within a country differs by size and (to a smaller degree) by sector. Within a country, the factor is usually highest for the smallest establishments, it is smallest for the largest establishments.

Table 6: Sampling rates, by country

Country	Establishments with 5+ employees in NACE Rev.2 A-S (in '000)	Net sample	Sampling rate (net interviews/universe)	Average factor (number of establishments represented by 1 interview, on average)
AT	134	1.503	0,011	89
BE*	115	1.506	0,013	76
BG	82	755	0,009	109
CH	176	1.502	0,009	117
CY*	14	757	0,054	18
CZ*	108	1.552	0,014	70
DE	1.206	2.264	0,002	533
DK	96	1.513	0,016	63
EE*	18	758	0,042	24
EL*	119	1.501	0,013	79
ES	458	2.266	0,005	202
FI	70	1.505	0,022	47
FR	682	2.251	0,003	303
HR*	42	740	0,018	57
HU*	109	1.504	0,014	72
IE(*)	68	1.999	0,029	34
IS*	6	753	0,126	8
IT	674	2.251	0,003	299
LT*	42	754	0,018	56
LU*	12	773	0,064	16
LV	33	756	0,023	44
MK*	18	752	0,042	24
MT*	8	453	0,057	18
NL	171	1.521	0,009	112
NO	102	1.951	0,019	52
PL	333	2.250	0,007	148
PT*	137	1.493	0,011	92
RO*	157	1.500	0,010	105
RS*	47	751	0,016	63
SE	143	1.512	0,011	95
SI*	21	1.067	0,051	20
SK*	59	756	0,013	78
UK	904	2.251	0,002	402
Total	6.365	45.420	0,007	140

* Universe (partly) estimated

The structure of the net sample

The targets for the size and sector structure agreed between EU-OSHA and Kantar in the fieldwork preparation phase could largely be met. There are however some exceptions to this:

- Some countries had problems to fill the quota for the smallest establishments (5 to 9), be it generally or only in selected sectors (e.g. NACE A or Q). This is partly related to weaknesses in the sampling frames (outdated addresses and altered workforce size) and partly to enhanced unit non-response rates for these small units.
- The situation in the largest size-class (250+) is diverse: Over all countries and sectors, the originally agreed targets for 250+ were not fully met, with 6% less interviews in the net sample than originally targeted. There are however also sectors where the targets were surpassed in this size-class.
- The overall sector targets have been met very well. Only in NACE A and C the achieved net samples are overall slightly lower than targeted, by 1 respectively 2 percentage points.

More details on the achieved size and sector structures can be found in chapter 7.4 of the Technical Report.

The magnitude of sampling errors in the full sample and in sub-samples

The variation inherent to any random sampling survey can be calculated and expressed by different measures, e.g. with the standard error or the confidence interval. The most suitable and illustrative measure for this survey is the confidence interval which therefore will be used here (see e.g. Eurostat 2009b, p.36).

The confidence interval shows within which range of values around the value estimated in the sample the true value lies with a certain probability, the most widely used level being the 95% probability level. At this probability level, the confidence interval can be directly calculated on base of the value of the standard error (se), by applying the formula $-1,96*se$ for the lower limit and $+1,96*se$ for the upper limit of the confidence interval.

The confidence interval (as well as the standard error) largely depends on the sample size – the larger the probability sample, the smaller the confidence interval and thus the sampling error is. But the sample size is not the only factor influencing the value of the confidence interval. It is also influenced by the total universe and by the distribution of answers, i.e. the percentage value measured for a variable: The further away a value is from both extremes (0% on the one end and 100% on the other), the larger the confidence interval is because the variance of answers is bigger in this case.

For the total net sample of 45.420 interviews, the confidence interval is very small for all statements based on unfiltered questions. Its maximum – not considering design effects – is $\pm 0,5$, meaning that the true value of a finding lies in a span defined by the measured percentage value $\pm 0,5\%$. The total number of $n = 45.420$ observations is thus large enough for quite precise statements on issues calculated over the whole sample (e.g. % of all 45.420 establishments doing regular risk assessments).

For sub-samples differentiated by either size, sector or country, the confidence interval is larger due to the reduced size of the sample taken into account:

- The 33 national net samples consist of 450 to 2.250 interviews.
- In the 19 sectors (NACE Rev.2 1-digit level), between 216 (NACE B=Mining and quarrying) and 8.508 interviews (NACE G = Wholesale and retail trade) were made.

- In the 4 size-classes, between 4.002 (250+ employees) and 19.082 (10-49 employees) interviews are available.

For the most detailed form of analysis, the analysis of the situation in establishments of a particular size in a specific sector within a single country, the available sub-samples can be very small and confidence intervals can reach values of around +/-10% or even more. Therefore, analyses on the level of single countries should be made in a differentiation either by size or by sector, but not by both at the same time. If analyses by sector AND size are considered important for the national level, they should be done only with very broadly summarized sectors (e.g. Producing Industries or Services) or in selected large NACE sections and countries.

The Table 7 below illustrates for two percentage values (80% and 50%) the decreasing of the confidence interval as the number of cases increases. For the illustration purpose we refrained from including design effect and finite population correction.

Table 7: Standard error (s.e.) and confidence interval (c.i.) at different levels of disaggregation

Size of the (sub-) sample (n =)	Percentage	Standard error s.e.	Confidence interval c.i. (95%)	Range	
				min	max
100	80,0%	4,0%	7,9%	72,1%	87,9%
450	80,0%	1,9%	3,7%	76,3%	83,7%
750	80,0%	1,5%	2,9%	77,1%	82,9%
1.000	80,0%	1,3%	2,5%	77,5%	82,5%
1.500	80,0%	1,0%	2,0%	78,0%	82,0%
2.250	80,0%	0,8%	1,7%	78,3%	81,7%
5.000	80,0%	0,6%	1,1%	78,9%	81,1%
20.000	80,0%	0,3%	0,6%	79,4%	80,6%
45.420	80,0%	0,2%	0,4%	79,6%	80,4%
100	50,0%	5,0%	9,8%	40,2%	59,8%
450	50,0%	2,4%	4,6%	45,4%	54,6%
750	50,0%	1,8%	3,6%	46,4%	53,6%
1.000	50,0%	1,6%	3,1%	46,9%	53,1%
1.500	50,0%	1,3%	2,5%	47,5%	52,5%
2.250	50,0%	1,1%	2,1%	47,9%	52,1%
5.000	50,0%	0,7%	1,4%	48,6%	51,4%
20.000	50,0%	0,4%	0,7%	49,3%	50,7%
45.420	50,0%	0,2%	0,5%	49,5%	50,5%

An example for interpretation:

For a question answered by 80% of respondents in a certain way (e.g. 80% of establishments stating to carry out risk assessments), the indicated range shows that due to sampling errors resulting from the chosen sample size, with a 95% probability²⁰ the true value lies between 76,3% and 83,7% for a sample of n = 450 interviews and between 78,3% and 81,7% for a sample of n=2.250 interviews²¹.

²⁰ The 95% probability means that when repeating the survey 100 times with the same number of interviews and an identical survey instrument, then in 95 of these 100 repetitions the value will probably lie in the indicated range. In 5 of the 100 repetitions, it is likely to lie outside this range.

²¹ This calculation takes only the sampling error into account. Differences between the measurement and the real occurrence of the phenomenon may additionally also arise from measurement errors (see next chapter).

For a question where the answers of the respondents vary more, with 50% answering in a certain way, the range is slightly larger. Here, the true value for a sample of $n = 450$ interviews lies between 45,4% and 54,6% and for a sample of 2.250 it lies between 47,9% and 52,1%, again with a 95% probability.

4.2 Errors of observation (measurement errors)

Measurement errors are “errors that occur during data collection and cause recorded values of variables to be different from the true ones (Eurostat 2009a, p.12)”.

According to the ESS standards, measurement errors can have different sources:

- a. Measurement errors deriving from the survey instrument
- b. Measurement errors due to erroneous data given by respondents
- c. Measurement errors related to the definition of respondents
- d. Measurement errors due to interviewer influences
- e. Measurement errors due to mode effects

In the following, the occurrence or dangers of these types of errors in the ESENER-3 survey will be analysed. Other than described in the ESS guidelines, section b (measurement due to erroneous data given by respondents) will be limited to consciously provided erroneous data. If a respondent unconsciously provides erroneous data, this is normally the result of a not totally clear questionnaire instrument. This type of error will therefore be dealt with under section a.

4.2.1 Measurement errors deriving from the survey instrument

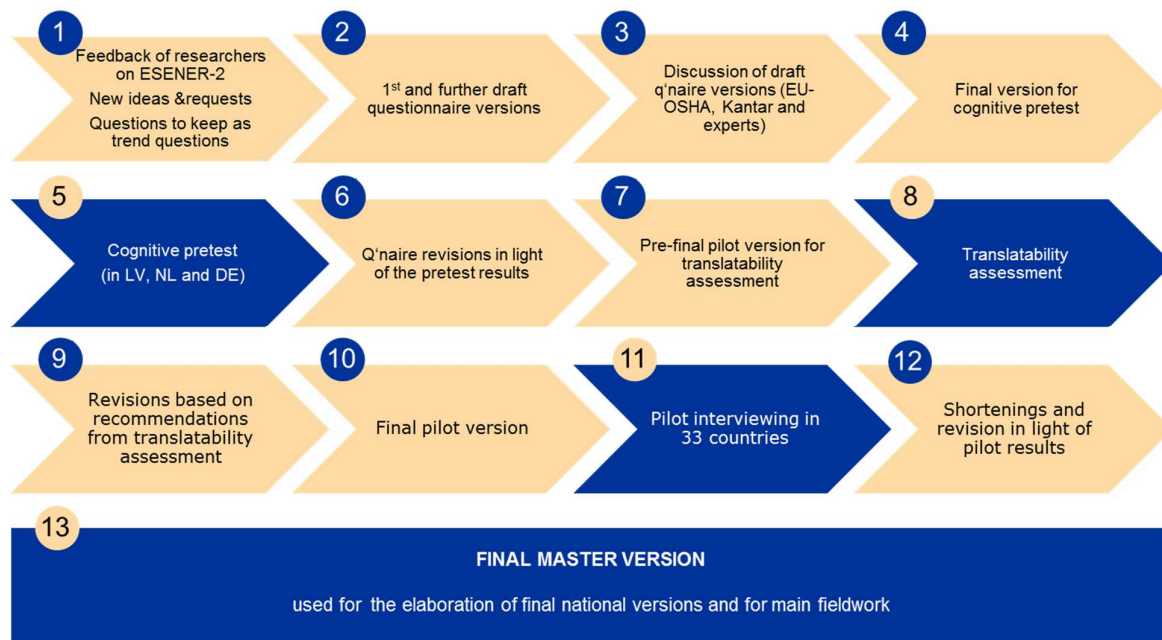
In the development of the survey instrument, the EU-OSHA project team, Kantar Public Division and experts from the Netherlands (TNO Netherlands) and Latvia (IOSEH at Stradins University Riga) were involved. Both the general survey concept and the questionnaire were intensely discussed in meetings, telephone conferences and written feedback rounds in order to minimize measurement errors deriving from the survey concept or the questionnaire. After each round of discussion, the questionnaire was revised, taking up the suggestions and annotations of the involved persons.

In addition, the survey instrument was extensively tested before being finalized and implemented for the full survey. In total, three test steps were carried out:

- (1) A cognitive pre-test with face-to-face interviews in 36 establishments in Latvia, the Netherlands and Germany (12 interviews per country). The main aims of this test were to analyse whether key questions and scales were well understood and interpreted in the intended way by respondents. Also, indicators for the relevance of the survey questions to respondents were collected.
- (2) A translatability assessment with rough translations of the questionnaire into three languages from different language groups, aiming at the identification of questions or formulations that may cause difficulties for translation, e.g. due to ambiguities. The languages included in this step were Swedish, French and Polish.

- (3) A pilot survey conducted in all 33 countries, with 30 or 40 interviews per country (40 in countries using more than one language version of the questionnaire). Main aims of this pilot were to test the survey infrastructure, the technical correctness of the CATI and CAWI scripts, the interview duration and the correctness and understandability of the national questionnaire versions.

Figure 1: Summary of the questionnaire development process



The various stages of questionnaire revision and testing helped to minimize measurement errors deriving from the survey concept and the questionnaire. In the development of the survey much care has been taken to formulate the questions as easy as possible, avoiding questions respondents might not be able to answer spontaneously. Additionally, in order not to urge respondents to give an answer if they are unsure about the situation, in all questions a “no answer” category was offered (though not read out). This further minimizes the number of erroneous answers. Not all respondents that are unsure about the correct answer do however use these categories – be it in order to please the interviewer or because they do not want to admit that they are not fully informed about the situation in their area of expertise within the establishment.

In spite of all preparatory measures, the occurrence of measurement errors originating in the survey instrument cannot be totally excluded. The feedback of interviewers from the different countries provides some hints on possible sources for remaining measurement errors arising from ambiguities in the questionnaire, but it is not possible to quantify these. The hints from the interviewers and supervisors were provided in detail to EU-OSHA in a separate documentation. The questions most frequently mentioned as unclear for part of the respondents were the following:

- Q151-1 to_-4: Here, the differentiation between some of these categories of OSH experts provided difficulties, especially the formulation “a generalist in health and safety”.
- Q262 and Q263: The answering scales used in these questions (major/minor/no reason respectively major/minor/no difficulty) proved to be unclear to respondents in some countries
- Q300-Q308: The section on psycho-social risks provided some difficulties particularly in small establishments

- Q310-312: The newly inserted questions on new technologies were partly difficult to understand, particularly for small establishments not using any more specific new technologies; specific items most often mentioned as problematic are Q310-3 (robots that interact with workers) and Q310-6 (wearable devices)

4.2.2 Measurement errors due to erroneous data given by respondents

In employer surveys, the phenomenon that questions are consciously answered erroneously can mainly be attributed to the desire to present the own organisation in a positive light (socially acceptable answers). This type of erroneous answers can be expected mainly in questions characterising the employer as a 'good' or 'bad' employer and particularly in questions concerning the compliance with legal obligations. In spite of the assurance about the anonymity of their answers given at the very beginning of the survey, in ESENER-3 such socially desirable answers may have occasionally occurred, particularly for questions referring to the implementation of health and safety provisions that are either obligatory (such as risk assessments or health and safety training for employees) or at least good practice. Socially desirable answers may have occurred particularly in the following questions:

Q150	Offer of regular medical examinations
Q151_1-4	Existence/availability of health and safety expertise in the establishment
Q155	Existence of a document on health and safety responsibilities
Q161	Existence of a procedure to support employees returning from long-term sick leave
Q250/Q257	Risk assessments and their documentation
Q263	Barriers for addressing health and safety issues
Q300-302	Action plans/procedures to deal with different types of psychosocial risks
Q350_3/_4	Formal employee representation in OSH issues (in countries/size-classes where this is obligatory)
Q355	Training of employees on health and safety issues

There is no reliable external source available to check whether social desirability is a major issue in the answers to these questions. The social desirability effect may vary between countries and between the different types of respondents (managers, employees in charge of health and safety etc.), depending e.g. on the stability of employment relations or the general trust in the confidentiality of survey data.

Some indications on possible measurement errors due to social desirability effects can be derived from a series of logistic regression analyses carried out on base of the ESENER-3 data-set. In these, it was checked whether the function of the respondent in the establishment (Q113) had - among other factors - an influence on the answers provided on core topics such as the performance of risk assessments, the training of employees and the set-up of an action plan on work-related stress.

In these models, some influence of the respondent type on the answers could be identified, but the influence is all in all not very strong for most aspects analysed. There are two exceptions: The respondents classifying themselves as health and safety specialists stated considerably more often than other respondent types that in their establishments workplace risk assessments are carried out and that ergonomic equipment has been provided in the past 3 years. This is however not necessarily a social desirability effect. On the contrary, the existence of an internal health and safety specialist can also be a sign that the establishment is very committed to health and safety.

4.2.3 Measurement errors related to the definition of respondents

Closely related to measurement errors deriving from the survey instrument and those owed to erroneous data provided by respondents are measurement errors related to the definition of respondents. For ESENER-3 (as already for ESENER-2), *“the person who knows best about health and safety in this establishment”* was defined as the person to be targeted for the interview. In order to help with the identification of the proper respondents, additional hints were provided in the questionnaire text, on who this person could be. These hints were filtered by size and sector (managing director or branch manager in smaller establishments, technical director or personnel manager in larger establishments of the producing industries, just personnel manager in larger service establishments).

According to the feedback from (some of) the field institutes, it was not always easy to get to the right person within the firm if the gatekeeper had no knowledge on how health and safety is managed in the establishment. The information provided in the questionnaire and in the survey manual helped however to get through to the right persons and the routing was described by most field institutes as very clearly structured. The rather low item non-response rates measured for ESENER-3 for almost all questions are indicators that the identification of the adequate respondent was largely successful.

The only respondent category not meant to be targeted, but only tolerated if otherwise no interview with the selected establishment would have been possible was Q113_7 “An external health and safety consultant” (see also Q006, a question inserted to persuade respondents to answer the questionnaire locally instead of referring it to their external OSH consultant)²². The share of respondents of this type can therefore be used to a certain extent as a quality indicator regarding the choice of the definition.

As Table 8 shows, all in all n = 157 interviews were carried out with such external OSH consultants. This is 0,3% of the total sample and thus a very small group. Relatively high absolute numbers of interviews with this group of respondents were made in Spain (31), Romania (13) and Serbia (11). In relative terms (as share of the total national sample), Serbia and Spain stand out with the highest share of interviews done with this type of respondent, though also from this perspective, external OSH experts are a marginal group in all countries, with the maximum share of 1,5% (Serbia) in any of the national samples. These country differences are not clearly attributable to differences in the respondent identification process during the interview but may also reflect national differences in the degree of outsourcing of health and safety duties. The indicator is therefore to be interpreted with caution.

²² Depending on the number of establishments such external experts are responsible for, they may well be in a position to answer the questions for a particular client establishment, provided they feel entitled to that. Some external experts, e.g. occupational doctors, have even an office in the client firms where they are present in regular intervals, e.g. on one specific day per week.

Table 8: Number and share of interviews done with external OSH consultants, by country

Country	Number of interviews	Share of interviews	Country	Number of interviews	Share of interviews
AT	4	0,3%	IT	6	0,3%
BE	6	0,4%	LT	9	1,2%
BG	1	0,1%	LU	1	0,1%
CH	2	0,1%	LV	4	0,5%
CY	1	0,1%	MK	1	0,1%
CZ	9	0,6%	MT	2	0,4%
DE	6	0,3%	NL	2	0,1%
DK	3	0,2%	NO	1	0,1%
EE	2	0,3%	PL	2	0,1%
EL	2	0,1%	PT	6	0,4%
ES	31	1,4%	RO	13	0,9%
FI	1	0,1%	RS	11	1,5%
FR	1	0,0%	SE	1	0,1%
HR	0	0,0%	SI	6	0,6%
HU	7	0,5%	SK	6	0,8%
IE	5	0,3%	UK	5	0,2%
IS	0	0,0%	ALL:	157	0,3%

Interviews with external consultants were made in all size-classes. In relative terms, they are more frequent in the larger size-classes than in the smaller ones.

Table 9: Number and share of interviews done with external OSH consultants, by size

Size-class	Number of interviews with external OSH consultants (Q113_7 = 1 respectively Q113 = 7)	Share of interviews with external OSH consultants (Q113_7 = 1 respectively Q113 = 7)
5 to 9 employees	36	0,2%
10 to 49 employees	48	0,3%
50 to 249 employees	38	0,5%
250 or more employee	35	0,9%
ALL	157	0,3%

Influence of the respondent type on the answers to the questionnaire

The most important question for an assessment of the rather flexible respondent definition chosen for ESENER-3 is whether the broader variety regarding the OSH background of respondents resulting from this definition actually leads to systematic differences in the answers. Against the background of full comparability of the survey results, ideally no such differences do occur.

In order to assess this question, we picked out a number of key variables from ESENER-3 as dependent variables and analysed in a multi-variate logit regression model whether the type of respondent had any systematic influence on the answers to these key questions. The dependent variables used in these models were:

Model 1:	Q250	Risk assessment carried out? (yes/no)
Model 2:	Q355_1	Employee training on proper use & adjustment of equipment
Model 3:	Q355_3	Employee training on psychosocial risks

Model 4:	Q311	Discussion of OSH impacts of new technologies
Model 5:	Q300	Existence of an action plan on work-related stress
Model 6:	Q159	Increase of sickness absence in last 3 years (outcome variable)
Model 7:	Q202_4	Provision of ergonomic equipment in last 3 years

All models are analysing the respondent behaviour with regard to factual questions. Models 1, 2, 3, 6 and 7 are models calculated for the entire net sample. For model 4, only establishments working with new technologies were included. Model 5 was limited to interviews from establishments with 20 or more employees because the question was filtered on size.

As independent variables, the following 13 variables were introduced in each of these models:

- Country
- Mode (CATI or CAWI)
- Size-class (Number of employees)
- Sector of activity (7 sector groups)
- Type of ownership (public or private)
- Type of respondent (Q113, as single-punch)
- Single-site vs. multi-site organisation ("more_est" = part of a multi-site organisation)
- Usage of OSH expertise (if any of Q151_3, _4 or _5 = "yes")
- Existence of a general employee representation (Works council or trade union, variable q16612)
- Existence of a specific OSH representation (H & S representative or Committee, variable q350_3/4)
- Number of traditional risks existing at the workplace (number of risks Q200_1 to _9 with "yes")
- Number of psycho-social risks existing at the workplace (number of risks Q201_1 to _5 with "yes")
- Visited by the labour inspectorate in last 3 years

The results of these regression analyses are documented in detail in the Annex. In summary, the influence of the type of respondent on the answers to the questions selected as dependent variables in the seven models can be described as follows:

- In model 1 (regular risk assessments), there are several significant correlations between the respondent type and the answers: Compared to owners, managing directors and other types of general workplace managers, specific managers with OSH duties and particularly internal or external OSH experts are more likely to indicate the performance of risk assessments. For the managers with OSH duties and the internal OSH experts this correlation is highly significant and for the latter, it is also quite strong, with an odds ratio of 1,935, meaning that this respondent type is almost twice as likely to indicate the performance of risk assessment than the owners.
- In models 2 and 3 (provision of training to employees), the type of the respondent has only a small influence on the answers.
- In model 4 (discussion of the impact of new technologies on health and safety), managers with OSH duties as well as OSH experts are slightly more likely than owners, managing directors or other types of general workplace managers to indicate that such risks are discussed.
- In model 5 (action plan on work-related stress), again managers with OSH duties and OSH experts are more likely to indicate the existence of such a plan than respondents from the reference group "owners, managing directors and other types of general workplace managers". This correlation is highly significant, but only moderately strong, with these two groups being by 26% and 27%, respectively, more likely to indicate the existence of such a plan.

- In model 6, OSH specialists, employee representatives and other employees in charge of OSH indicate less often than owner or general managers that the sickness absence in the establishment has increased over the last 3 years. A reason for this difference might however be that these groups of people are less well informed about the sickness rates than the managers.
- In model 7, all other groups of respondents (except for the very few external OSH consultants) indicate more often than the owners and managing directors that the establishment has provided ergonomic equipment to employees.

While the models have shown that there are some correlations between the respondent type and the analysed OSH measures, it is not possible to clearly interpret these. It is e.g. not possible to say that another type of respondent would have provided a different answer for the same establishment.

4.2.4 Measurement errors due to interviewer influences

The survey was carried out by Kantar institutes or partner institutes of the Kantar network. With the sole exception of Ireland, which was fielded from the United Kingdom, the survey was conducted by experienced²³ local interviewers in each country. In order to minimize interviewer influences, all interviewers working on the project received a detailed briefing about the specific challenges of this survey before starting to interview. The briefings were held personally, by the local supervisors and/or fieldwork managers. In addition, written material on key issues and on particular questions was provided to interviewers in their local language as a back-up. The supervisors and local fieldwork managers in charge of the training had previously participated in two training seminars: A 2-day seminar held before the pilot survey and another 2-day training seminar held after the pilot and shortly before the launch of fieldwork for the main survey.

For this project, each national institute selected their most successful and experienced interviewers because surveys among organisations are generally among the biggest challenges for interviewers, particularly in the contact phase. These interviewers had previously received various general trainings and were constantly supervised by their CATI supervisors. To this end, supervisors regularly listened live²⁴ into at least 10% of interviews and contact attempts. The listening in is documented in detailed log files made available to EU-OSHA. The files documents the outcome of the listening-in for each interview and the number and share of interviews controlled from each single interviewer.

By way of the listening-in, any undesired interviewer influences could easily be detected and further individual or general training measures were immediately taken or the respective interviewer was removed from the ESENER-3 team. A few interviews were also disqualified in this context and removed from the data-set.

The number of interviewers working on this project in each country was relatively small. It was a pre-requisite of EU-OSHA to keep teams rather small and stable in order to ensure best performance. All in all, 820 interviewers were working on the project. On the average, thus $45.420/820 = 55$ interviews were done per interviewer. The number of different interviewers appointed to the project in each country was small enough to develop routine and expertise in this specific questionnaire. On the other hand, the interviewer teams were usually large enough as to ensure that any individual interviewer effect that

²³ This does hold only partly for the interviewers in Slovakia. Due to very low response rates and a very slow progress of fieldwork from the beginning, at the end numerous additional interviewers had to be used for completing the envisaged net sample size. These interviewers were not always experienced and extensively trained.

²⁴ In a few countries, the listening-in was partly or entirely made on base of the recorded interviews, i.e. after an interview had been made. In the concerned countries, all respondents were asked explicitly for their agreement to the recordings.

might persist in spite of the careful selection and supervision does not have an overly influence on the overall data²⁵.

To sum up, the careful selection of well experienced interviewers for the project, various training measures and a consistent control of the interviewer teams ensure that interviewer effects are kept to a minimum.

4.2.5 Effects of the mode of data collection

In an effort to check the data-set for possible influences originating from the different modes (CATI and CAWI) applied for data-collection, a variable for the mode of data collection was included among the independent variables in the multi-variate regression models described in chapter 4.2.3 and annexed to this report.

The results of the regression models show that:

- For models 1 to 4, some mode effects can be observed in the data: Respondents of the online interview report less often than CATI respondents that in their establishment risk assessments are done and employees trained on issues such as the adjustment of their workstation or on psycho-social risks. They also report less often that possible health and safety impacts of new technologies used in the establishment have been discussed. For all four measures, the differences are highly significant, but only moderate, with online respondents being between 34% (training on the adjustment and use of the workstation) and 16% (training on the prevention of psycho-social risks) less likely to report these measures.
- The mode has in turn no statistically significant effect on the respondent's answer to the question whether or not an action plan on work-related stress exists in the establishment or whether ergonomic equipment has been provided in the last 3 years. Also no significant effects were identified on the outcome indicator measuring changes in the level of sickness absence over the last 3 years.

Table 10: Summary of multi-variate regression analyses on mode effects in ESENER-3

Analysis No.	Dependent variable	Significance level	Direction of the correlation: + more likely to do/have this in the establishment - : Less likely to do/have this in the establishment	Odds ratio
1	Workplace risk assessments carried out regularly (yes)	***		0.797
2	Training on proper use/adjustment of workstation (yes)	***	-	0.658
3	Training on prevention of psycho-social risks (yes)	***	-	0.839
4	Discussion of impact of new technologies on OSH (yes)	***	-	0.814
5	Existence of an action plan on work-related stress (yes)	-		
6	Outcome indicator: Increased sickness absence (yes)	-		
7	Provision of ergonomic equipment in last 3 years (yes)	-		

To sum up, it can be observed that overall, the mode of interviewing seems to have moderate effects on some of the health and safety measures tested in the model. In all cases where effects can be observed, the online respondents indicate a lower OSH performance. Whether this reflects differences

²⁵ In a few cases, interviewers did a quite large number and/or share of the interviews. These were however very experienced and successful b2b interviewers and it was decided together with EU-OSHA not to put strict limits regarding the maximum number of interviews an interviewer may do. The constant listening-in even to the interviews of these experienced interviewers and their broad experience make undesired interviewer effects very unlikely.

in the (self-)selection (all online respondents had refused the CATI interview previously and might be representing establishments that are in tendency less committed to OSH) or in the way how the survey is conducted (with online interviewing being more anonymous, not involving an interviewer when answering the questionnaire) cannot be determined on base of the available data.

In view of the only moderate differences between respondents of both modes and the fact that in most countries only a minority of interviews has been conducted online, the mode does generally not present a big limitation regarding the comparability of the data.

This holds however only partly for the data of the Czech Republic where about one third of all interviews was conducted online. Here, the increased share of online interviews may have more influence on the comparability across countries and over time (trend developments).²⁶

For all three measures, the mode regression models had shown a significantly lower affirmation rate for the CAWI interviews (see Table 13). A repetition of these calculations just for the Czech data (eliminating the data from the other countries and eliminating “country” among the controlled variables) shows however quite significant mode effects for model 1 (risk assessment) and model 2 (training on the adjustment and proper use of the workplace equipment): According to this, the probability that respondents of the online version in the Czech Republic indicated the performance of risk assessment is by about 48% lower than for the CATI respondents. For the training on the workstation usage and adjustment, it is even by 52% lower. In ESENER-3, 73,8% of establishments in the Czech Republic had indicated to regularly carry out risk assessments, 63,1% indicated to do this type of training.

Table 11: Mode effects in the Czech subsample

Analysis No.	Dependent variable	Significance level	Reference	
			Direction of the correlation: + more likely to do/have this in the establishment - : Less likely to do/have this in the establishment	Odds ratio
1	Workplace risk assessments carried out regularly (yes)	***	-	0.518
2	Training on proper use/adjustment of workstation (yes)	***	-	0.480
3	Training on prevention of psycho-social risks (yes)		-	
4	Discussion of impact of new technologies on OSH (yes)		-	
5	Existence of an action plan on work-related stress (yes)	-		
6	Outcome indicator: Increased sickness absence (yes)	-		
7	Provision of ergonomic equipment in last 3 years (yes)	-		

As regards comparability, this means that compared to the Czech ESENER-2 data and to other countries in ESENER-3, the Czech ESENER-3 data are in tendency too “pessimistic”, at least in regard with risk assessments and the training on the workplace usage and adjustment. In tendency, it can however also be assumed that the online data reflect the reality better than the CATI data – be it because the offer of this additional answering option helped to get interviews from potential non-respondents or because some survey questions were answered more honestly in absence of an interviewer.

Table 12 below shows that the quite strong mode effect on risk assessments shown for the Czech Republic cannot be generalized: None of the four further countries with a relatively high number of online interviews (>100 accepted interviews) shows significant mode effects with regard to the workplace risk assessments. In Denmark, there were no significant mode effects in any of the seven models.

²⁶ Among the 4 measures for which mode effects were identified in the regression models 1-7, there are 3 measures asked about in an identical way in ESENER-2 and -3. The trend developments for these measures do not show any very obvious impact of the mode on the comparability in this bivariate perspective: While risk assessments dropped by 3 percentage points (from 77% to 74%) and the training on psycho-social risks by 1 percentage point, training on the usage of equipment was reported more often (+3 percentage points) in ESENER-3.

Table 12: Mode effects in the subsamples for DK, ES, LU and NL (countries <100 online int.)

Analysis No.	Dependent variable	Reference: CATI mode							
		DK		ES		LU		NL	
		Significance level	Odds ratio	Significance level	Odds ratio	Significance level	Odds ratio	Significance level	Odds ratio
1	Workplace risk assessments carried out regularly (yes)	-		-		-		-	
2	Training on proper use/adjustment of workstation (yes)	-		-		***	0,451	-	
3	Training on prevention of psycho-social risks (yes)	-		-		-		-	
4	Discussion of impact of new technologies on OSH (yes)	-		-		***	0,549	***	0,425
5	Existence of an action plan on work-related stress (yes)	-		**	0,409	-		-	
6	Outcome indicator: Increased sickness absence (yes)	-		-		-		-	
7	Provision of ergonomic equipment in last 3 years (yes)	-		-		-		-	

4.3 Processing errors

4.3.1 Data cleaning and coding errors

In computer assisted surveys, posterior data cleaning and editing is hardly necessary. Filters are set by the program so that the interviewers or respondents cannot make any filtering mistakes. Likewise, data provided in different questions that are somehow related to each other can be checked on the spot for their plausibility by programmed plausibility checks with the option for respondents to correct the contradictory data on the spot. The main area of application of such plausibility checks are questions asking for numerical data. The main questionnaire for ESENER-3 does however hardly contain such data – the only numerical data is the number of employees asked in Q102. Plausibility checks were therefore not integrated into ESENER-3.

For ESENER-3, various general data checks carried out with the CATI and CAWI scripts before the launch of fieldwork and weekly data checks with a programmed syntax after the launch of fieldwork ensured that filtering mistakes were minimized.

There was no filtering mistake identified for the main survey. In the first data-check and tabulations done after the launch of fieldwork, it was however decided to revisit the filtering rules for two questions:

- In Q164a, asking whether the respondent him-/herself has received any training on how to manage health and safety, the filter was broadened: Respondent of the type Q113_4 = 1 (health and safety officers) were additionally included in the base for this question. They had been excluded because it was at some instance considered as self-evident that health and safety officers would be trained for this job. But another check of the ESENER-2 data (where this group of respondents had been included in the question on training) revealed that 9% of those attributing themselves to the category “health and safety officer” indicated not to have received any such training. For the (few) cases done with the old filtering, interviews with this type of respondents were set to “No answer” in Q164a.
- In Q254, in turn, a closer look at the data suggested that the filter in this question should be further narrowed down. The question asks whether risk assessments also cover workplaces outside the premises of the establishment (other than home office workplaces). It was initially asked to all respondents indicating to carry out risk assessments (Q250 = 1). Here, an additional filter “and if Q107 = 1” was introduced, now excluding those who at the beginning of the interview indicated not to have any workplaces outside the premises (home office workplaces were not considered here as they were asked about in the previous question Q106). In interviews from

establishments not having any such external workplaces Q254 was set to “missing” so that the question has the same base for all interviews in the data-set.

A specific type of coding applied in ESENER-3 was the ex-post coding of the sector descriptions. These sector descriptions were available for all interviews where respondents considered the sector code from the sampling frame as wrong in Q108 and could also not clearly identify the appropriate sector with the search engine in Q109. The coding was done by a small team of professionals (two persons) specialized in the sector coding. This method of sector coding was chosen in order to ensure a very high quality of the coding, with a minimization of coding mistakes and a very high degree of harmonization and coherence.

4.3.2 Weighting errors

The weighting procedure was carried out with great care, treating all countries according to uniform rules while at the same time specificities of the national net samples such as cells with a low number of interviews were taken into account by individual treatments (summarizing of cells). The weighted samples were controlled by comparing the weighted structures (size and sector) with the respective structures of the national universe. The results of this comparison were made available to EU-OSHA. Further aspects of the weighting, including measures on the effectiveness of the weighting, are set out in the Technical Report (chapter 9) and are not repeated here. Errors from the weighting process as such are unlikely to have occurred. Each step of the weighting was controlled by a second statistician.

But as described more in detail in the Technical Report, in many countries statistics on the distribution of establishments (the relevant sampling unit and unit of enquiry) were not available. In view of this situation, estimates on the distribution of establishments were made, based on company statistics and (in particular segments of the universe) data from the Labour Force Survey. Though great care was taken in the elaboration of these best estimates, there is always the possibility of remaining estimation errors. Estimates necessarily have to be derived from some type of proxy data and based on assumptions, which may not always be fully correct.

In absence of reference data for the countries requiring estimates, it is hardly possible to exactly verify the estimates. They can only be assessed with regard to their plausibility, e.g. in how far the estimated universe size and structure is in line with that of other countries (with reliable official statistics) that are similar in key criteria such as their size, their geographical location or their economic history and situation. A particular difficulty for the estimates done for ESENER-3 is that in many cases it was difficult to identify such “reference countries” because e.g. most of the Eastern and Central European states do not have any establishment statistics available (exceptions are Latvia and Bulgaria).

The countries where no genuine establishment-level statistical data were available for the weighting are therefore likely to have a higher degree of weighting errors than the data from countries where adequate official statistics were available for the weighting.

5 Timeliness and punctuality

The ESS standard for quality reports defines timeliness of statistical outputs as *“the length of time between the event or phenomenon they describe and their availability (Eurostat 2009a, p.19)”*.

Punctuality is defined as *“the time lag between the release date of data and the target date on which they were scheduled for release as announced (...) or previously agreed among partners (Eurostat 2009a, p.19)”*.

In the original project planning agreed with EU-OSHA at the project’s kick-off meeting on 20th June 2018, dates were fixed for the key survey steps. These dates were later broken down to a finer time planning considering each working step more in detail. For each step, the scheduled date and the final execution date were recorded. For some of these finer broken down working steps, some days of delay occurred.

The two key dates set in the time-planning were the launch of fieldwork for the main survey, scheduled for 1st April 2019 and the date for the delivery of the integrated, checked and weighted data-set and cross-tabulations, scheduled for 27th September 2019.

In spite of smaller delays in some of the preparatory steps, in 21 of the 33 countries fieldwork started on exactly the scheduled date. In the other countries, the start was also only slightly later, with 8 countries starting with the first live interviews on the subsequent day, 2nd April, 2 countries starting on the 3rd April and the final two countries (CY and MT) starting on 4th April, that means 3 days later than scheduled. The later start in some countries was not owed to the survey preparation – samples and scripts were ready for all countries by 1st April – but rather to local holidays (1st April was a public holiday in Cyprus) or local capacity planning.

The criterion “punctuality” can be fully assessed only at a later stage, after publication of all reports planned on the ESENER-3 results. A first publication in form of a 15-page brochure over the highlight findings, has however already been published by EU-OSHA on its website on 15 November 2019, i.e. about 3,5 months after finalization of fieldwork and just 7 weeks after the delivery of the cleaned and weighted data-set. The brochure is available for free download under <https://osha.europa.eu/en/publications/third-european-survey-enterprises-new-and-emerging-risks-esener-3/view>.

Table 13: Time periods for key project steps (plan, reality and comments)

Key working steps	Comments	Planned start	Planned finalisation	Actual finalisation
2018				
Inception meeting	The inception meeting marked the official start of the project work	20.6	21.6	21.6
Signature of contract by both parties	Date of signature (reference point for timelines on report deliveries)	22.6	22.6	22.6
Delivery of Inception Report	Postponed due to prioritization of other working steps; revised time planning sent on 21/06/2018 (together with presentation of kick-off meeting); detailed minutes as a key element of the Inception Report relevant for the further working steps delivered already by 03/07/2018,	5.7	5.7	21.9
Questionnaire development meeting (between EU-OSHA, Kantar and experts)	Took place as scheduled	12.7	13.7	13.7
Development of a first draft questionnaire version	A first draft questionnaire was drafted after the first questionnaire meeting between EU-OSHA, Kantar and the contracted experts from TNO and IOSEH; slight delay of 1 working day	25.6	13.7	19.7
Finalisation of the cognitive test instruments (English master questionnaire, interview guidelines)	Delay of 3 working days in the finalisation of the test instruments, mainly because discussions on questions took longer than anticipated	27.7	27.7	1.8
Fieldwork for cognitive pre-test	Fieldwork for the cognitive pre-test took 1 week longer than scheduled, but this slight delay was compensated by a quicker finalisation of the pre-test analysis and reporting which were already started during fieldwork for the cognitive pre-test. The delay in the interviews was caused by holiday absences of researchers (LV, Berlin) and respondents. The FW period for the cognitive tests could be scheduled only after the kick-off meeting for the project when holiday plans on part of researchers had already been made.	3.8	31.8	7.9
Delivery of the report on the outcomes of the cognitive pre-test	Delay of one working day in the delivery of the report in order to be able to fully integrate the findings from the latest interviews; postponement actively offered by EU-OSHA since a delivery on 10/09 was considered sufficient for a good preparation on the meeting for discussing the pre-test results.	7.9	7.9	10.9
Meeting between Kantar, EU-OSHA and experts to discuss the findings from the cognitive pretest	The meeting took place as scheduled, the slight delays in the finalisation of interviews and report could thus be caught up without any negative impact on results etc.	13.9	14.9	14.9
Revision of questionnaire and finalisation of master version for translatability check	Final changes to the questionnaire as scheduled.	17.9	21.9	21.9
Translatability assessment	Results of translatability check received on 28 September from translators; passed on to EU-OSHA on the next working day, 1 October.	24.9	28.9	1.10
Revision of the questionnaire in light of the findings from the translatability assessment	A few days more than anticipated were needed for the final round of modifications to the questionnaire, based on the translatability assessment and other input.	1.10	4.10	11.10.
OK of EU-OSHA for the final master version for the pilot		5.10	5.10	11.10
Preparation of translation master (integrating existing trend translations) and set-up of translation tools	Translation master set up, differentiating between 3 types of questions (trend, modified trend, new).	8.10	12.10	11.10
Preparation of a brief explaining key terms for translation	Translators notes drafted and agreed with EU-OSHA.	8.10.	12.10.	11.10
Pilot questionnaire sent to translators T1 and T2 for new translations and check of trend questions	Master questionnaire distributed 2 days after the translator training.	12.10.	12.10.	12.10
Collection of detailed up-to-date statistical information on the distribution of establishments and employees by the national fieldwork institutes and from Eurostat	Collection of statistical information from countries launched on 16 October, i.e. with 1 week delay compared to the original planning. Request for LFS statistics sent to Eurostat on 18 October. By 23 October, a majority of countries had sent us their statistics, some counts were however still pending. We have been receiving further statistics still in December - for some specific counts from their national statistical offices, countries had to wait several weeks. In these cases, however, best estimates had been made for the sample orders, based on the universe structures used in ESENER-2. Any incoming new figures will be used for an update of the universe statistics. The updated figure will be used for drawing the initial samples and for the weighting of the data after finalisation of fieldwork.	9.10	19.10.	23.10.

Key working steps	Comments	Planned start	Planned finalisation	Actual finalisation
2018				
Webex training for translators and verifiers (2 or more sessions)	Training for the translators took place on 10 October. The training for verifiers/adjudicators is scheduled for 26 October, 11 a.m.	15.10	19.10	26.10
Verified national versions sent to EU-OSHA for checks by domain experts	First batch for Focal Point checks sent on 12 November, second batch on 13 November, i.e. with 1-2 working days delay.	09.11.	09.11.	13.11.
Processing and estimation of figures on the universe of establishments (and employees in establishments) and elaboration of national sampling plans	This very work intense process took longer than planned; we were however aware that the planning of this step was from the beginning very ambitious, in an effort to avoid having FW done in the summer holiday period.	22.10	16.11	04.12.
Elaboration of national sampling plans and agreement with EU-OSHA on the plans	Agreement with EU-OSHA on the principles for the sampling plans (2:1 establishment/employee proportionality). Sampling plans were then sent out without prior explicit agreement on individual plans.	29.10.	23.11	04.12.
Kick-off seminar with national project managers and supervisors	Seminar held at the date agreed for this in late summer	22.11	23.11	23.11
Finalisation of English version of training documents for interviewers (pilot) and other auxiliary material for the pilot (e.g.motivation letter)	Interviewer instructions finalised as part of the first interim report delivered on 30/10. Further auxiliary material finalised on 3/12 and subsequently sent for translation.	30.11	30.11	03.12.
Acquisition and compilation of national sampling frames by local fieldwork partners	The launch date for the acquisition of the samples had to be postponed. Sampling plans were finalised by 4 December and subsequently sent to the national fieldwork partners. As agreed at the occasion of the inception meeting with national fieldwork partners, countries that could not provide their sample in time before the launch of the pilot phase were provided with a centrally sourced sample from a commercial provider.	20.11	7.12	8.3
Finalisation of translation process for the pilot survey, including integration of feedback from domain experts and final proofreading	A first batch of 12 national questionnaire versions was delivered on 14 December. The questionnaire translation process for the other countries was not yet finalised by that date. A major reason for this are delays in the feedback of the Focal Point experts. In case of shared languages, these delays have repercussions also on the finalisation of other language versions (e.g. Italian, German and French due to the delayed feedback from CH). All language versions could be finalised before the Christmas holiday period, but there was a last query (on the health and safety representation for Estonia) that was finally clarified in the beginning of January only.	14.12	14.12	9.1
Translation of training documents for interviewers (pilot) and other auxiliary material for the pilot (e.g. motivation letter)	All ancillary materials were sent out for translation and most versions were finalised by 14 December. The only ancillary material still not translated at that date was the data privacy sheet. This was however also translated by 21.12.	3.12	21.12	21.12
Check and "correction" of national gross samples	Check of the first gross samples launched on 11 December. The process took longer for some countries which had difficulties to get the samples within reasonable time; the process of receiving new (partial) samples and checking these all in all took until 25 March.	3.12	12.1	25.3
Programming and testing of the international master pilot script	The international CATI master script was finalised by 17/12, including the Irish boost questions. The CAWI master script was finalised by 21/12.	15.10	11.1	21.12
Programming and testing of the international master pilot script and all national script versions	The overwriting and testing of national script versions started on 18/12. It was finalised for most countries by 11 January, the last national version was finalised and tested by 16/01.	15.10	11.1	16.1
Installation and test of questionnaires and address management on the TripleC server	A first test version for upload on the TripleC server was provided on 13/12 to enable a timely preparation of the technical environment. Subsequently, further adaptations were made and the interface between the sample management system was further improved.	9.1	9.1	13.12
Start of fieldwork (FW) for pilot survey	Postponement of the start of main fieldwork by 3 working days from 14/01 to 17/01 agreed with EU-OSHA shortly before the Christmas break; the main reason for the postponement were: * Delays in the translation process for some countries, mostly due to late feedback from part of the respective national Focal Point experts * More time needed for the over-writing Finally, FW was launched in 2 countries on 14/01 (IT and NL). In most other countries, FW was launched on 17/01. Luxembourg launched fieldwork on 21/01 only, i.e. with 2 working days delay compared to the main FW start and 5 days delay as compared to the originally envisaged FW start. The reason were issues with the switch language functions that had to be solved before FW start.	14.1	14.1	17.1

Key working steps	Comments	Planned start	Planned finalisation	Actual date
2019				
End of fieldwork (FW) for pilot survey	The end of FW for the pilot survey has been postponed by 2 working days, from Friday 1st February to Tuesday 5th February. A shortening of the fieldwork period in order to cope with the 3 days delay from the start of FW would have been possible, but would have reduced the chances to collect experiences with the telephone reminder calls for respondents not having completed the online interview sent to them.	1.2	1.2	5.2
Compilation and delivery of pilot report	The planned delivery date for the pilot report was postponed by 2 working days, from Friday 08th February to Tuesday 12th February, in accordance with the postponement of the pilot fieldwork end date.	4.2	8.2	12.2.
Decisions on changes to the survey instrument and accompanying materials, based on the pilot results	Due to the slight delays in the pilot fieldwork and the reporting on the pilot results, the decision-making on final changes was accordingly also delayed by a few days.	20.2	20.2	27.2.
Translation of final changes to the questionnaire and fieldwork materials	The modified master questionnaire was sent to the translation team on 27th February. Translations were received back on 6th March. After some checks, they were distributed to the adjudicators on 7th March. The adjudicated versions were then distributed to the central coordination team in 4 batches between 8th and 12th March.	21.2	8.3	12.3.
Pre-fieldwork seminar with national fieldwork managers and supervisors	The pre-fieldwork seminar, the 2nd among the two ESENER-3 seminars, took place as scheduled on 18th and 19th March 2019.	18.03.	19.03.	19.3
Drawing initial samples for the main survey	For some countries, the compilation of the full gross sample (covering all sectors and sizes) took longer than planned. The drawing of the samples itself then could be done quickly within two days, on 26th and 27th of March. In a few countries, however, some addresses were still missing and needed to be added after the drawing of the initial samples.	18.2	15.3	27.3.
Launch of the main survey (part of Work Package B)	In spite of the slight delays in the start of the pilot fieldwork and the subsequent steps (delivery of the pilot report and decisions on the final changes), in the majority of countries (21 of 33) fieldwork was launched at the date envisaged in the original planning of June 2018. The remaining countries mostly started with the training on 1st April, but with live interviewing only on the following day, 2nd April 2019. The latest two countries to start FW were CY and MT, doing the first interviews on 4th April, i.e. 3 days after the envisaged start.	1.4.	1.4	1.4
Delivery of final fieldwork material, including the national language questionnaires			26.4.	
Delivery of a first set of cross-tabulations with interim data, for checking purposes			8.4.	08.4.
Weekly updates on the fieldwork progress, starting 1 week after the launch of FW	Starting from 8th April, weekly updates on the fieldwork progress were provided every Monday, based on the state of FW from Friday end of the day		8.4.	8.4.
Extraction of the final data, editing, coding and weighting of the final data set	The establishment-proportional weighting was finalised at the targeted date (27.09.); For the employee-proportional weighting, 2 additional working days were needed due to the need for adjustments of outliers	06.08.	27.9	1.10.
Delivery of the weighted and integrated data-set	The weighted and integrated data-set was delivered as scheduled, on 27.09. As for the employee-proportional weighting factor, an update was however made on 2nd October only		27.9.	27.9.
Delivery of cross-tabulations by country, by sector and by size	The cross-tabulations were delivered with a slight delay. The establishment-weighted tables were delivered on 30 September (+1 working day), the employee proportional tables were delivered on 2nd October (+3 working days)		27.9	2.10.
Delivery of the Technical Report	The first version of the full Technical Report was delivered on time; the final version taking into account all feedback of the EU-OSHA team was then delivered on 29 November 2019		31.10.	30.10.
Delivery of the Quality Report and any other agreed survey documentation	The delivery of all further document was scheduled for CW 50 the latest. The Quality Report and all remaining further documentation will actually be delivered by 5th December (CW 49)		13.12	6.12

6 Coherence and comparability

The coherence and comparability of statistical outputs are key criteria for any cross-national survey. The ESS standard defines coherence as referring to *“the degree to which the statistical processes by which they [the statistical data] were generated used the same concepts – classifications, definitions, and target populations - and harmonised methods. Coherent statistical outputs have the potential to be validly combined and used jointly (Eurostat 2009a, p.21)”*. Comparability is defined as a *“special case of coherence”, referring to cases where “the statistical outputs refer to the same data items and the aim of combining them is to make comparisons over time, or across regions, or across other domains (Eurostat 2009a, p.21)”*.

Achieving a high degree of coherence and comparability is a task that put challenges in different phases of the development and implementation of the survey, from the development of the questionnaire(s) to the process of data processing. In the following, the main aspects contributing to a high coherence and comparability are described.

6.1 Development of the survey concept

In the development phase of the survey, the concepts to be used were discussed between EU-OSHA, Kantar Public Division and (for some aspects) the international expert group. The international composition of this team and its experience in cross-country research ensured that the survey concept was shaped in a way which made it applicable to all European countries. Moreover, the main survey characteristics had already undergone a large-scale empirical “test” in form of the ESENER-2 conducted in 2014 in 36 countries. Most key aspects of the survey design were adapted from ESENER-2 with no changes (e.g. the definition of respondents, the introduction to the survey) or with only moderate modifications (e.g. screening procedure, online interview reminders). The basic concept of the survey had thus been tested already on a very large scale in 2014.

For ESENER-3, apart from issues related to the questionnaire approach, a number of more technical aspects can be considered as prerequisites for the coherence of such a survey:

A common sampling unit and unit of enquiry in all countries

Surveys among organisations can be addressed either to the unit of enterprises/companies or to establishments/local units. In the case of the series of ESENER surveys, a decision on the establishment/local unit level had been taken by EU-OSHA already before the launch of the first survey wave. The consequent application of this level to all countries is a pre-requisite for achieving a high coherence: Therefore, it was important to use the same unit of enquiry for sampling in all involved countries. In 16 of the 33 countries included in the survey, this was achieved by drawing the samples directly from an address register based on local units/establishments. For the other 17 countries, such a register did not exist. Therefore, a specific screening procedure was applied there in order to get a selection of addresses of local units from an enterprise-based register.

The usage of genuine establishment samples is clearly preferable to the application of the screening procedure to company-level addresses. But in view of the lack of such registers in many countries, the applied screening procedure can be considered as an efficient way to get a sample that is reasonable representative as regards the inclusion of subsidiaries from multi-site organisations.

The screening procedure actually helped to get a good mixture of multi- and single-site organisations and within the latter a mixture of headquarters and subsidiaries in the final sample. The procedure was however less successful than in ESENER-2, with a considerably lower number of interviews being

conducted with addresses of subsidiaries newly taken up during the contact with the first address (usually the headquarters).

In consequence, the number and share of interviews conducted with multi-site organisations and – among these - the number and share of interviews conducted with subsidiaries is quite low in several screening countries. Countries with a particularly low share of interviews in multi-site organisations are Bulgaria (29 interviews, 4% of the total sample, unweighted), Romania (67 interviews, 4%) and Slovakia (43 interviews, 6%)²⁷. These shares improve for most countries in the weighted sample (see Table 15). But though the real share of multi-sites respectively subsidiaries is not known, multi-sites and subsidiaries are likely to remain underrepresented in the sample for these and other screening countries.

A major reason for the practical difficulties in a successful application of the screening was a declining willingness to provide personal data such as the names and contact details of potential respondents in the subsidiaries since the new GDPR (European General Data Protection Regulation) entered into force in May 2018²⁸. The GDPR notably raised the awareness on such data protection issues. Another factor that might have led to the relatively poor success of the screening for some countries is the increased complexity of the modified screening procedure in ESENER-3, with considerably more details about the company structure being asked from respondents.

²⁷ See Table 14 for the unweighted figures and Table 15 for the weighted figures. In both tables, the screening countries are in red letters.

²⁸ https://ec.europa.eu/info/priorities/justice-and-fundamental-rights/data-protection/2018-reform-eu-data-protection-rules/eu-data-protection-rules_en

Table 14: Number and share of multi-site organisations and of subsidiaries, by country, unweighted

Country		Interviews in total	Single-site organisation	Multi-site organisation	NA/DK type of multi-site org.	Hereof head-quarters	Hereof subsidiaries	NA type of multi-site org.
at	n	1.503	1.028	473	2	332	134	7
	in %	100%	68%	31%	0%	70%	28%	1%
be	n	1506	989	515	2	299	215	1
	in %	100%	66%	34%	0%	58%	42%	0%
bg	n	755	725	29	1	20	8	1
	in %	100%	96%	4%	0%	69%	28%	3%
ch	n	1.502	939	562	1	263	295	4
	in %	100%	63%	37%	0%	47%	52%	1%
cy	n	757	675	80	2	67	12	1
	in %	100%	89%	11%	0%	84%	15%	1%
cz	n	1.552	1.318	227	7	146	66	15
	in %	100%	85%	15%	0%	64%	29%	7%
de	n	2264	1337	924	3	509	403	12
	in %	100%	59%	41%	0%	55%	44%	1%
dk	n	1.513	645	859	9	260	594	5
	in %	100%	43%	57%	1%	30%	69%	1%
ee	n	758	641	115	2	81	34	0
	in %	100%	85%	15%	0%	70%	30%	0%
el	n	1.501	1.264	235	2	172	62	1
	in %	100%	84%	16%	0%	73%	26%	0%
es	n	2.266	1.391	874	1	564	298	12
	in %	100%	61%	39%	0%	65%	34%	1%
fi	n	1.505	856	646	3	253	385	8
	in %	100%	57%	43%	0%	39%	60%	1%
fr	n	2.251	1.325	909	17	235	623	51
	in %	100%	59%	40%	1%	26%	69%	6%
hr	n	740	551	188	1	151	36	1
	in %	100%	74%	25%	0%	80%	19%	1%
hu	n	1.504	1.339	160	5	113	44	3
	in %	100%	89%	11%	0%	71%	28%	2%
ie	n	1.999	1.133	863	3	218	631	14
	in %	100%	57%	43%	0%	25%	73%	2%
is	n	753	502	250	1	158	90	2
	in %	100%	67%	33%	0%	63%	36%	1%
it	n	2.251	1.699	552	0	351	199	2
	in %	100%	75%	25%	0%	64%	36%	0%
lt	n	754	666	86	2	60	25	1
	in %	100%	88%	11%	0%	70%	29%	1%
lu	n	773	583	188	2	103	81	4
	in %	100%	75%	24%	0%	55%	43%	2%
lv	n	756	636	119	1	98	21	0
	in %	100%	84%	16%	0%	82%	18%	0%
mk	n	752	625	126	1	90	36	0
	in %	100%	83%	17%	0%	71%	29%	0%
mt	n	453	377	76	0	57	18	1
	in %	100%	83%	17%	0%	75%	24%	1%
nl	n	1.521	918	597	6	281	285	31
	in %	100%	60%	39%	0%	47%	48%	5%
no	n	1.951	1.007	934	10	188	734	12
	in %	100%	52%	48%	1%	20%	79%	1%
pl	n	2.250	1.590	656	4	333	322	1
	in %	100%	71%	29%	0%	51%	49%	0%
pt	n	1.493	1.185	305	3	258	45	2
	in %	100%	79%	20%	0%	85%	15%	1%
ro	n	1.500	1.429	67	4	42	23	2
	in %	100%	95%	4%	0%	63%	34%	3%
rs	n	751	666	84	1	68	16	0
	in %	100%	89%	11%	0%	81%	19%	0%
se	n	1.512	750	760	2	288	457	15
	in %	100%	50%	50%	0%	38%	60%	2%
si	n	1.067	867	200	0	164	35	1
	in %	100%	81%	19%	0%	82%	18%	1%
sk	n	756	709	43	4	24	16	3
	in %	100%	94%	6%	1%	56%	37%	7%
uk	n	2.251	1.090	1.158	3	347	793	18
	in %	100%	48%	51%	0%	30%	68%	2%
TOTAL	n	45.420	31.455	13.860	105	6.593	7.036	231
	in %	100%	69%	31%	0%	48%	51%	2%

Table 15: Share of multi-site organisations and subsidiaries, by country, establ-prop. weighted

Country	Single-site organisation	Multi-site organisation	NA type of organisation	Hereof head-quarters	Hereof subsidiaries	NA type of multi-site org.
at	76%	24%	0%	72%	27%	1%
be	70%	30%	0%	55%	45%	0%
bg	94%	6%	0%	38%	60%	2%
ch	66%	34%	0%	40%	59%	1%
cy	90%	10%	0%	67%	32%	1%
cz	82%	18%	0%	50%	45%	5%
de	70%	30%	0%	50%	49%	1%
dk	48%	51%	1%	26%	74%	0%
ee	84%	16%	0%	53%	47%	0%
el	80%	20%	0%	43%	57%	0%
es	69%	31%	0%	59%	39%	1%
fi	60%	40%	0%	33%	66%	1%
fr	62%	37%	1%	24%	71%	5%
hr	75%	25%	0%	62%	37%	1%
hu	88%	12%	0%	54%	45%	1%
ie	58%	42%	0%	25%	73%	2%
is	58%	42%	0%	36%	64%	1%
it	79%	21%	0%	60%	40%	0%
lt	87%	13%	0%	45%	54%	0%
lu	78%	21%	0%	41%	57%	2%
lv	84%	16%	0%	67%	33%	0%
mk	81%	18%	0%	53%	47%	0%
mt	83%	17%	0%	73%	26%	1%
nl	69%	31%	0%	39%	55%	6%
no	55%	45%	0%	17%	81%	1%
pl	76%	24%	0%	46%	54%	0%
pt	82%	18%	0%	71%	29%	0%
ro	96%	4%	0%	43%	55%	3%
rs	91%	9%	0%	79%	21%	0%
se	53%	47%	0%	35%	63%	2%
si	80%	20%	0%	60%	40%	0%
sk	92%	8%	0%	28%	69%	3%
uk	53%	46%	0%	24%	74%	1%
All	69%	30%	0%	40%	59%	2%

Usage of international classifications

The usage of international classifications is an important prerequisite to establish international comparability of a survey. In ESENER-3, international classifications were relevant for the sector of activity only. The application of the European NACE Rev. 2 sector classification for the sampling of establishments ensured that in all countries exactly the same sectors were covered.

Some of the address sources used for the survey (e.g. the Dun & Bradstreet register used for the United Kingdom and Bill Moss for Ireland) were originally not classified by NACE, but used the SIC codification instead. SIC is a similar classification of activities and outside Europe it is even more common than the NACE classification. In the preparation of the sampling for ESENER-3, all addresses from registers using the SIC codes were first classified into NACE Rev.2 codes by way of a transformation key. For all relevant survey steps, these addresses were thus also available in the NACE classification. The share of addresses considered as not correctly classified (on the NACE Rev.2 2-digit level) by respondents in Q108 was however somewhat higher than average for Ireland and the United Kingdom.

This inaccuracy in the transformation owed to the slightly restricted compatibility of the two classification systems was not a very problematic issue since respondents were asked to correct the sector description from the sampling frame if it was considered wrong. The data-set includes the corrected NACE codes.

Usage of uniform size bands

The usage of the same size-bands in all countries and for all relevant survey steps (definition of targets, sampling, fieldwork monitoring and steering, weighting) was another measure ensuring coherence of the data. Some countries could not deliver the statistical information in the desired size differentiation. In these cases, it was necessary to estimate the distribution for the requested size-bands so that in all countries, finally exactly the same size-bands could be used for the definition of targets, the monitoring of fieldwork and the weighting.

Likewise, the sample frame providers in a few countries delivered addresses in slightly deviating size-bands (e.g. 200+ in Finland instead of 250+ as foreseen). For these countries, in the sample preparation phase estimates were made on how many units would fall into the size fraction that is different (e.g. in this case the number/share of addresses in size-class 200-249). The initial samples were then adapted to these estimates before being ordered and drawn. By this way, the deviations did not lead to any sample disproportionalities.

6.2 Conceptualisation and organisation of the survey

ESENER-3 (as ESENER-2) was to a large degree coordinated centrally in order to ensure a high coherence and comparability in all aspects of the survey (both concepts are closely interrelated). The philosophy was however not to do everything centrally, but to decentralize the steps and decisions that can better be taken at the local level. Main steps that were done centrally were:

- The translation of the questionnaire and the motivation letter by translators 1 and 2 (with a questionnaire and language expert from each local fieldwork partner being in charge of adjudicating the two translations²⁹),
- The programming of the questionnaire scripts (with a chance to do additional tests on the local level and to propose changes, e.g. in terms of the script layout)
- The definition of the targets for the sampling structure (with a chance for local fieldwork partners to intervene if targets were considered as totally unrealistic)
- Check of provided samples and drawing of gross samples (from the sampling frames selected and acquired by the local fieldwork partners on base of shared instructions)
- The entire data management, including data checks and data cleaning
- Progress monitoring and monitoring of net sample structures
- Sample management and sample releases (in coordination with the countries)
- Central training of the local fieldwork managers and CATI supervisors (as “multipliers” in charge of passing the relevant instructions on to the local interviewers).

The central coordination of these steps facilitated the execution of the survey in exactly the same way in all involved countries. Fieldwork itself was done at the local level by national partner institutes. This included the following steps that were all done locally:

²⁹ Following the TRAPD method: Translation, Review, Adjudication, Pre-testing and Documentation

- Selection of interviewers for the survey (with the selection being centrally checked on criteria such as the team size and interviewer's level of experience)
- Training of interviewers (based on the instructions provided centrally to the local fieldwork managers and supervisors and on the written instructions made available in all local languages)
- Checks on the documentation of the local quality control measures related to the performance of interviewers
- Reporting on changes in the interviewer teams and on specific fieldwork challenges

Some crucial steps such as training and instruction as well as the control and monitoring of fieldwork were done on both levels, i.e. centrally and locally, with a partly different focus (e.g. local control of the work of individual interviewers, central control over the fieldwork progress and the development of non-response).

6.3 Usage of harmonized national questionnaire versions

An essential pre-requisite for a coherent and comparable production of survey data is the use of fully harmonized national questionnaire versions. This implies two aspects: Firstly, the elaboration of a master questionnaire equally applicable in all countries. And secondly, the elaboration of fully harmonized national language questionnaire versions on base of the master questionnaire.

A common master questionnaire for all countries

In order to achieve a high degree of coherence of the data collected in the different countries, a common master questionnaire version for use in each participating country was developed. National variations of the questions were mostly not necessary since the questions are applicable to the situation in all 33 countries. This holds particularly for the EU countries where a number of EU-Directives on health and safety at the workplace are valid in all countries and are meant to ensure a certain degree of standardization regarding the health and safety measures to be taken.

The only exception where national deviations within the master questionnaire were foreseen is question Q350_1 to _4 asking for different forms of employee representation. Here, specific national forms of representation at the workplace may be set up and some of the forms mapped in Q350 do not exist in all of the countries. In Germany and Austria, for example, the employee representation at the workplace level is generally with works councils while in other countries (e.g. Cyprus) it is with so called "shopfloor" trade union representations. In the majority of countries, however, both forms of representation do exist. Such national differences are taken into account by the filtering in Q350 – forms of representation not relevant in a country were not asked there.

A specific issue for ESENER-3 are the national boost questions. All participating countries had been offered the opportunity to order additional interviews, so-called sample boosts. These had to be paid from national means and came on top of the interviews ordered and paid via EU-OSHA. In ESENER-3, all countries with sample boosts were offered the opportunity to include 3 or 4 own, country specific questions on top of the international master questionnaire set up by EU-OSHA. Three countries – Ireland, Norway and Slovenia – made use of this opportunity by ordering sample boosts and including additional questions not asked in the other countries. These questions were placed at the end of the topical section where they belonged to (e.g. risk assessments) or at the very end of the questionnaire. By this way, it was ensured that they would not influence the answers on any following questions and thus, would not endanger the full cross-national comparability of the core questions.

The additional boost questions included in ESENER-3 are marked with "bo_" in the variable labels. The following are the boost questions:

Table 16: List of additional questions asked in boost countries only

Variable	Countries	Content
Q201_6bo	Norway and Slovenia	Additional psycho-social risks
Q201_7bo	Norway and Slovenia	Additional psycho-social risks
Q201_8bo	Norway and Slovenia	Additional psycho-social risks
Q252_7bo	Ireland	Additional risks evaluated in workplace risk assessments
Q252_8bo	Ireland	Additional risks evaluated in workplace risk assessments
Q252_9bo	Ireland	Additional risks evaluated in workplace risk assessments
Q359bo	Norway and Slovenia	Employment policy and procedures regarding workers with disabilities
Q360bo	Norway and Slovenia	Policy and procedures regarding older workers
Q361bo	Norway and Slovenia	Particular attention paid to hazards of younger workers
Q362bo	Ireland	Policy on ergonomig risks
Q363bo	Ireland	Accreditation of the establishment with the Irish body INAB
Q364	Ireland	Benefits from the INAB accreditation

Elaboration of national questionnaire versions – the translation process

After finalisation of the English language master questionnaire, translation into different languages started. For the translation process, Kantar Public Division in Munich cooperated with the translation team of Kantar Public Division in Brussels. The translation team in Brussels was built up in 2005 for the translation of the Eurobarometer questionnaires and has since then been responsible for the translation of numerous Eurobarometer, for ESENER-1 (2009) and for other demanding cross-national surveys with high quality demands.

The entire translation process was handled by the translation team in Brussels, with additional supervision and support by the project team at Kantar Public Division in Munich.

For all national translations, the TRAPD team translation model was applied. TRAPD (Translation, Review, Adjudication, Pretesting, and Documentation) is currently widely considered as best practice approach to survey translations (Harkness et al. 2010: 128ff.).

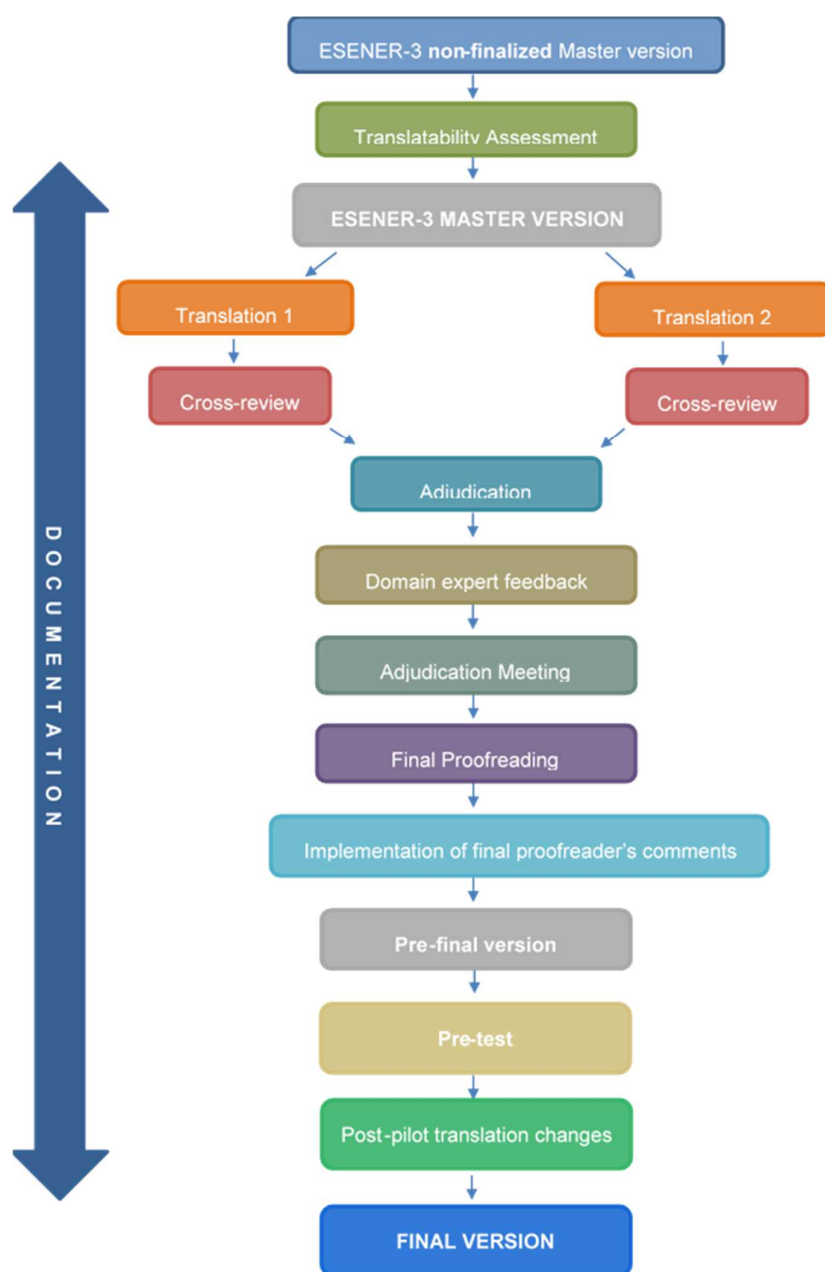
In the following, the main steps of this process as applied to ESENER-3 are shortly summarized (the process is documented in more detail in the Translation Report):

- A few weeks before the launch of the translation process, a translatability assessment was carried out for all new or modified questions in order to identify formulations that might be difficult to translate into other languages or might cause ambiguities in some languages. To this end, translators from three languages of different language families (Polish as Slavic language, French as a Latin language, Swedish as Germanic/Nordic language) elaborated rough translations of the master version and documented any difficulties and potential ambiguities encountered hereby. The results were analysed by linguists at the translation unit and proposals for modifications on the master questionnaire were made. Where considered useful by the questionnaire development team, these proposals were then implemented.
- In a first step of the genuine translation process, then two independent translations were produced by different translators of the translator pool (T1 and T2) of Kantar Public Division in Brussels.

- The two translations were reviewed by an adjudicator. As adjudicators, survey experts from the local fieldwork partners were used. These were native speakers of their local language and had additionally an excellent command of the English language. The adjudicators compared the two translations and selected the best of the translations for each question. If both translations for a question or an item were considered as not optimal, the adjudicator proposed a third, new version. The result of this step was a translation that, from the adjudicator's point of view, represented the optimum. In this version, all decisions made by the adjudicator were documented and justified in order to allow retracing and discussing the decisions later on in the further translation process.
- These versions were then sent to national health and safety experts of the Focal Point network of EU-OSHA (domain experts) where they were checked again, with a particular focus of the correctness of all specific health and safety terminology used in the questionnaire and on the new and modified questions. Trend questions were meant to be annotated only in case the version used in ESENER-2 should be considered as really wrong or misleading.
- The comments and changes proposed by the Focal Point experts were read and annotated by both translators (T1 and T2) and the adjudicator. For all language versions, then review meetings with the participation of the adjudicator and both translators were held, mostly in form of Webex meetings. In these review meetings, the decisions of the adjudicator and the proposals of the Focal Point expert were discussed question by question. Focal Point expert proposals were hereby accepted in most but not all cases. Proposals not in line with the intentions of the master questionnaire were rejected. All decisions were again documented. The project coordination team at Kantar Public Division in Munich was consulted during this process on all decisions where the local teams were not totally sure. Where necessary, the project team at EU_OSHA was additionally consulted.
- After the adjudication meeting, the adjudicator finalized the implementation and annotation of all decisions commonly taken. This version was then sent for a final proofreading step. In the final proofreading done by another translator, it was checked that all relevant questions and items from the master were completely translated and that no grammar, spelling or other mistakes remained in the final version.
- The national language versions resulting from this were used for the pilot survey conducted in all countries. After the pilot, feedback from the local fieldwork partners (interviewers, supervisors and local fieldwork managers) on the national questionnaire versions was collected and analysed, in addition to analyses made on base of the data. Where necessary, revisions were discussed and implemented. These revisions then were subject to a final check by the adjudicator responsible for the language version.

This last step led to the final national language versions that were used in the main survey. During survey fieldwork, no changes to any of the national language versions were made. The entire process including all decisions and changes on individual questions or items has been documented for each country in an Excel file including separate columns for each translation and checking step.

Figure 2: Summary of the translation process



6.4 Comparability of the data – indications from the data-set

In many countries, ESENER is the only survey among organisations on the topic of health and safety at work. Only few countries (e.g. LV, DK and DE) have conducted own national representative surveys on that topic in recent years. In those countries that actually did conduct such surveys, hardly any variable is directly comparable because questions were usually asked in a different way. It is therefore for most countries not possible to directly compare the survey results with data from other sources for the purpose of a verification of the data.

A comparison of the ESENER-3 data with the data of the previous wave ESENER-2 shows that for most questions asked in an identical way results are well comparable on the aggregated level (across all countries), with values either remaining quite stable or showing plausible developments (e.g. increased awareness for some OSH aspects). On the national level, however, for some of the trend variables the measurements change considerably more.

7 Respondent burden

The burden for respondents can be calculated on base of the figures on the interview duration and the number of completed interviews. On the average of all countries, the survey took 24 minutes for completion. This is calculated by summing up the average duration from all countries and dividing this by the number of countries (33), not considering the different national sample sizes. For a calculation of the respondent burden it is thus more precise to calculate the respondent burden individually for each country and to then sum this up. This way the respondent burden of the survey results in a total close to 18,000 hours. Assuming an 8 hour working day as standard, this is equivalent to 2.240 person days.

Table 17: Respondent burden in terms of interviewing time, by country

Country	Duration (in minutes)	Interviews	Total time burden (minutes)	Total time burden (hours)
AT	23,2	1.503	34.794	580
BE	28,2	1.506	42.499	708
BG	28,9	755	21.820	364
CH	26,1	1.502	39.232	654
CY	21,7	757	16.389	273
CZ	27,3	1.552	42.323	705
DE	23,4	2.264	52.932	882
DK	25,1	1.513	37.931	632
EE	23,0	758	17.434	291
EL	18,3	1.501	27.468	458
ES	23,5	2.266	53.296	888
FI	28,9	1.505	43.540	726
FR	22,4	2.251	50.355	839
HR	28,4	740	21.038	351
HU	21,8	1.504	32.787	546
IE	21,3	1.999	42.619	710
IS	24,4	753	18.336	306
IT	17,8	2.251	40.113	669
LT	22,0	754	16.588	276
LU	27,8	773	21.489	358
LV	25,3	756	19.104	318
MK	22,1	752	16.597	277
MT	24,2	453	10.976	183
NL	23,6	1.521	35.865	598
NO	23,2	1.951	45.263	754
PL	25,1	2.250	56.520	942
PT	19,2	1.493	28.695	478
RO	27,5	1.500	41.175	686
RS	23,5	751	17.626	294
SE	28,3	1.512	42.714	712
SI	25,2	1.067	26.835	447
SK	23,8	756	17.978	300
UK	19,0	2.251	42.656	711
ALL	24,0	45.420	1.074.989	17.916

To this interviewing time, some time for the first contact(s) within the organisations (for the identification of the right respondent, for scheduling an interviewing time, but also for telephone attendance in case of refusals) have to be added for those cases where the interview did not take place immediately at the first contact attempt. For this preparatory time, measurements are not available. The time invested by establishments where finally no interview took place (due to refusals or other reasons) is not included in these calculations either.

On the part of an individual participant, the burden for a completed interview is with about 24 minutes of average interviewing time at the limit of what in many countries is considered as acceptable for a telephone interview among businesses. Any substantial further prolongation of the survey instrument is not recommended since this would lead to a serious quality loss in the answers.

Though the survey is not particularly difficult as regards the scales used (mostly simple yes/no scales), it is demanding due to its numerous question batteries which can make the interview at times a bit fatiguing for the respondent. Any overstretching of respondent's capacities to concentrate in the telephone interview goes at the expense of data quality, particularly towards the end of the interview.

8 Confidentiality, transparency and security

Full confidentiality of the data is a legal prescription and was guaranteed to interviewees at the beginning of the interview. To ensure full confidentiality, addresses and data were immediately separated. By 31 December 2019 the latest, all addresses will be deleted. Exceptions to this rule are the interviews from respondents who explicitly agreed to be re-contacted for a possible follow-up study. For these, additional address information was taken up during the interview and will be stored for a longer time. These addresses will be deleted by 31 December 2021 the latest.

The allowance of a re-contact was asked from respondents in a set of two questions:

Q401:

May we or another research institute contracted by the European Agency for Safety and Health at Work contact you again later if we should have additional questions for a follow-up study based on your answers in this survey?

Q402 (if agreement to Q453):

In order to re-contact you for this purpose, can I ask your name, email address and direct phone number please?

In countries where these questions were considered as not sufficient to comply with the national data protection legislation, additional information on the address storage and the possible re-contacting was provided in the interview.

All respondents were ensured that any analyses would be done in a generalized way, not allowing conclusions on any specific firms. The strict separation of addresses and data is a prerequisite for this guarantee, but it is not necessarily sufficient because even the anonymised data may in some cases contain enough information as to clearly identify a particular firm on this base. This is particularly the case for interviews made in sections of the economy in which only very few establishments exist in the universe (e.g. large firms of small sectors of activity in small economies).

To inhibit an ex-post identification of individual firms on base of the interview data, the following further measures were taken:

- The data on the size of the actual workforce (Q102) is not made available in numerical terms, but only as roughly summarized categories.
- Though data were collected with a sector differentiation at the NACE Rev.2 2-digit level, this differentiation will not be included in the data-set made available to any external researchers. There, only the NACE Rev.2 1-digit level is made available (with the small sectors NACE B,D and E grouped to sector group BDE and the sectors K and L grouped to sector group KL).
- Likewise, data collected on the distribution of establishments by region (within a country) was not included in the data-set.

9 Annexes

Logit regression models for analysis of mode and respondent effects

Table 18: Model 1: Dependent variable Q250 (Risk Assessments)

Explaining factors: (NA = No answer)	Level of significance: * low (p<0,05) ** medium (p<0,01) *** high (p<0,005)	Direction of the correlation: + : more likely to carry out risk assess. - : less likely to carry out risk assess.	Degree/extent of the correlation (odds ratio)
Interview Mode (Reference: CATI)			
CAWI	***	-	0,797
Country (Reference: United Kingdom)			
Austria	***	-	0,161
Belgium	***	-	0,240
Bulgaria	-		
Croatia	***	-	0,576
Cyprus	***	-	0,094
Czech Republic	***	-	0,344
Denmark	-		
Estonia	***	-	0,400
Finland	***	-	0,509
France	***	-	0,228
Germany	***	-	0,232
Greece	***	-	0,151
Hungary	***	-	0,594
Iceland	***	-	0,131
Ireland	***	-	0,506
Italy	***	+	1,553
Latvia	-		
Lithuania	***	-	0,186
Luxembourg	***	-	0,095
Macedonia	***	-	0,131
Malta	***	-	0,325
Netherlands	***	-	0,553
Norway	***	-	0,556
Poland	***	-	0,590
Portugal	***	-	0,498
Romania	-		
Serbia	*	-	0,705
Slovakia	***	-	0,210
Slovenia	-		
Spain	***	+	1,600
Sweden	*	-	0,760
Switzerland	***	-	0,091
Size class (Reference: size_1 = 5 to 9 employees)			
size_2 (10 to 49 employees)	***	+	1,277
size_3 (50 to 249 employees)	***	+	1,908
size_4 (250 or more employees)	***	+	2,855
Sector group (Reference: Sector_4 = NACE G,H,I,R)			
NACE A	*	+	1,291
NACE B, D, E, F	***	+	1,581
NACE C	***	+	1,428
NACE J, K, L, M, N, S	***	-	0,823
NACE O	***	-	0,714
NACE P, Q	***	+	1,234

Table continued

Explaining factors: (NA = No answer)	Level of significance: * low (p<0,05) ** medium (p<0,01) *** high (p<0,005)	Direction of the correlation: + : more likely to carry out risk assess. - : less likely to carry out risk assess.	Degree/extent of the correlation (odds ratio)
Function of the respondent (Q113; Reference: Owner, managing director)			
Manager without OSH duties	**	-	0,930
Manager with OSH duties	***	+	1,270
OSH specialist	***	+	1,935
Employee representative in charge of OSH	-		
Another employee in charge of OSH	*	-	0,904
External OSH consultant	*	+	2,244
Ownership type (Q111; Reference: Not public)			
Public organisation	***	-	0,863
NA ownership type	-		
Type of organisation (Q100; Reference: Single-site organisation)			
Part of a multi-site organisation	***	+	1,310
OSH experts (Q151_3 to _5; Reference: None of these OSH expert types available in the establishment)			
OSH expertise available	***	+	2,525
Economic situation (Q400; Reference: Very good economic situation)			
Quite good economic situation	-		
Economic situation neither good nor bad	***	-	0,851
Quite bad economic situation	***	-	0,780
Very bad economic situation	-		
NA economic situation	***	-	0,738
Employee representation in terms of OSH (Q350_1 to 4; Reference: non-existence of the respective position/body of representation)			
General ER (works council or trade union)	***	+	1,268
Health and safety representation (OSH representative or co	***	+	2,064
Risk profile (Reference: 0 risks of the respective type; numerical; odds ratio per risk)			
"Traditional" health and safety risks (Q200_1 to _10)	***	+	1,127
Psychosocial risks (Q201_1 to _7)	***	-	0,894
Visit from labour inspectorate (Q154; Reference: not visited in last 3 years)			
Visited	***	+	1,590
NA Labour inspectorate visit	***	-	0,788
Dependent variable: Workplace risk assessments regularly carried out ("yes" to Q250)			

Table 19: Model 2: Dependent variable Q356_1: Employee training the workstation

Explaining factors: (NA = No answer)	Level of significance: * low (p<0,05) ** medium (p<0,01) *** high (p<0,005)	Direction of the correlation: + : more likely to provide such training - : less likely to provide such	Degree/extent of the correlation (odds ratio)
Interview Mode (Reference: CATI)			
CAWI	***	-	0,658
Country (Reference: United Kingdom)			
Austria	***	-	0,509
Belgium	***	-	0,354
Bulgaria	***	-	0,410
Croatia	***	-	0,496
Cyprus	***	-	0,449
Czech Republic	***	-	0,340
Denmark	***	-	0,524
Estonia	***	-	0,422
Finland	***	-	0,698
France	***	-	0,164
Germany	***	-	0,450
Greece	***	-	0,621
Hungary	***	-	0,453
Iceland	***	-	0,498
Ireland	***	+	1,492
Italy	-		
Latvia	***	-	0,392
Lithuania	***	-	0,282
Luxembourg	***	-	0,293
Macedonia	***	-	0,423
Malta	***	-	0,578
Netherlands	***	-	0,285
Norway	-		
Poland	***	-	0,220
Portugal	***	-	0,605
Romania	***	-	0,214
Serbia	**	-	0,733
Slovakia	***	-	0,461
Slovenia	-		
Spain	-		
Sweden	***	-	0,567
Switzerland	***	-	0,484
Size class (Reference: size_1 = 5 to 9 employees)			
size_2 (10 to 49 employees)	*	+	1,057
size_3 (50 to 249 employees)	**	+	1,112
size_4 (250 or more employees)	***	+	1,463
Sector group (Reference: Sector_4 = NACE G,H,I,R)			
NACE A	-		
NACE B, D, E, F	-		
NACE C	**	+	1,108
NACE J, K, L, M, N, S	-		
NACE O	-		
NACE P, Q	***	-	0,745

Table continued

Explaining factors: (NA = No answer)	Level of significance: * low (p<0,05) ** medium (p<0,01) *** high (p<0,005)	Direction of the correlation: + : more likely to provide such training - : less likely to provide such	Degree/extent of the correlation (odds ratio)
Function of the respondent (Q113; Reference: Owner, managing director)			
Manager without OSH duties	**	-	0,917
Manager with OSH duties	-		
OSH specialist	-		
Employee representative in charge of OSH	*	-	0,895
Another employee in charge of OSH	***	-	0,849
External OSH consultant	-		
Ownership type (Q111; Reference: Not public)			
Public organisation	***	-	0,854
NA ownership type	-		
Type of organisation (Q100; Reference: Single-site organisation)			
Part of a multi-site organisation	-		
OSH experts (Q151; Reference: None of the mentioned OSH expert types available in the establishment)			
OSH expertise available	***	+	1,828
Economic situation (Q400; Reference: Very good economic situation)			
Quite good economic situation	*	-	0,935
Economic situation neither good nor bad	***	-	0,821
Quite bad economic situation	***	-	0,648
Very bad economic situation	*	-	0,751
NA economic situation	***	-	0,763
Employee representation in terms of OSH (Q350_1 to 4; Reference: non-existence of the respective position/body of representation)			
General ER (works council or trade union)			
Health and safety representation (OSH representative or co	***	+	2,064
Risk profile (Reference: 0 risks of the respective type; numerical; odds ratio per risk)			
Risk_tard (Q200_1 to _9)	***	+	1,127
Risk_psycho (Q201_1 to _7)	***	-	0,960
Visit from labour inspectorate (Q154; Reference: not visited in last 3 years)			
Visited	***	+	1,221
NA Labour inspectorate visit	-		
Dependent variable: Employee training on the use and adjustment of their working equipment and furniture ("yes" to Q355_1)			

Table 20: Modell 3: Dependent variable Q355_3: Training on prevention of PSR (Q355_3)

Explaining factors: (NA = No answer)	Level of significance: * low (p<0,05) ** medium (p<0,01) *** high (p<0,005)	Direction of the correlation: + : more likely to provide such training - : less likely to provide such training	Degree/extent of the correlation (odds ratio)
Interview Mode (Reference: CATI)			
CAWI	***	-	0,839
Country (Reference: United Kingdom)			
Austria	***	-	0,445
Belgium	***	-	0,471
Bulgaria	***	-	0,410
Croatia	***	-	0,404
Cyprus	***	-	0,442
Czech Republic	***	-	0,290
Denmark	***	-	0,573
Estonia	***	-	0,466
Finland	***	-	0,693
France	***	-	0,480
Germany	***	-	0,371
Greece	***	-	0,689
Hungary	***	-	0,642
Iceland	**	-	0,775
Ireland	*	-	0,862
Italy	-		
Latvia	*	-	0,817
Lithuania	***	-	0,498
Luxembourg	***	-	0,349
Macedonia	***	-	0,571
Malta	***	-	0,503
Netherlands	***	-	0,458
Norway	***	-	0,773
Poland	***	-	0,805
Portugal	***	-	0,654
Romania	***	-	0,664
Serbia	***	-	0,768
Slovakia	***	-	0,397
Slovenia	-		
Spain	-		
Sweden	***	-	0,776
Switzerland	***	-	0,471
Size class (Reference: size_1 = 5 to 9 employees)			
size_2 (10 to 49 employees)	-		
size_3 (50 to 249 employees)	-		
size_4 (250 or more employees)	***	+	1,398
Sector group (Reference: Sector_4 = NACE G,H,I,R)			
NACE A	***	-	0,799
NACE B, D, E, F	***	-	0,703
NACE C	***	-	0,768
NACE J, K, L, M, N, S	***	+	1,161
NACE O	***	+	1,179
NACE P, Q	***	+	2,154

Table continued

Explaining factors: (NA = No answer)	Level of significance: * low (p<0,05) ** medium (p<0,01) *** high (p<0,005)	Direction of the correlation: + : more likely to provide such training - : less likely to provide such training	Degree/extent of the correlation (odds ratio)
Function of the respondent (Q113; Reference: Owner, managing director)			
Manager without OSH duties	-		
Manager with OSH duties	-		
OSH specialist	-		
Employee representative in charge of OSH	-		
Another employee in charge of OSH	***	-	0,897
External OSH consultant	-		
Ownership type (Q111; Reference: Not public)			
Public organisation	***	+	1,109
NA ownership type	-	-	0,998
Type of organisation (Q100; Reference: Single-site organisation)			
Part of a multi-site organisation	***	+	1,251
OSH experts (Q151; Reference: None of the mentioned OSH expert types available in the establishment)			
OSH expertise available	***	+	1,645
Economic situation (Q400; Reference: Very good economic situation)			
Quite good economic situation	***	-	0,839
Economic situation neither good nor bad	***	-	0,712
Quite bad economic situation	***	-	0,643
Very bad economic situation	***	-	0,520
NA economic situation	***	-	0,747
Employee representation in terms of OSH (Q350_1 to 4; Reference: non-existence of the respective position/body of representation)			
General ER (works council or trade union)			
Health and safety representation (OSH representative or co	***	+	2,064
Risk profile (Reference: 0 risks of the respective type; numerical; odds ratio per risk)			
Risk_tard (Q200_1 to _9)	-		
Risk_psycho (Q201_1 to _7)	***	+	1,044
Visit from labour inspectorate (Q154; Reference: not visited in last 3 years)			
Visited	***	+	1,218
NA Labour inspectorate visit	*	+	1,118
Dependent variable: Training on how to prevent psychosocial risks such as stress or bullying ("yes" to Q355_3)			

Table 21: Model 4: Dependent variable Q311: Discussion of OSH impacts of new technologies

Explaining factors: (NA = No answer)	Level of significance: * low (p<0,05) ** medium (p<0,01) *** high (p<0,005)	Direction of the correlation: + : more likely to discuss OSH impacts - : less likely to discuss OSH impacts	Degree/extent of the correlation (odds ratio)
Interview Mode (Reference: CATI)			
CAWI	***	-	0,814
Country (Reference: United Kingdom)			
Austria	***	-	0,362
Belgium	***	-	0,637
Bulgaria	***	-	0,738
Croatia	***	-	0,367
Cyprus	***	-	0,356
Czech Republic	***	-	0,670
Denmark	***	-	0,487
Estonia	***	-	0,490
Finland	***	-	0,344
France	***	-	0,428
Germany	***	-	0,442
Greece	***	-	0,339
Hungary	***	+	3,078
Iceland	***	-	0,394
Ireland	*	-	0,842
Italy	***	-	0,378
Latvia	**	+	1,287
Lithuania	***	-	0,293
Luxembourg	***	-	0,396
Macedonia	*	-	0,782
Malta	*	-	0,768
Netherlands	*	-	0,840
Norway	***	-	0,392
Poland	***	-	0,647
Portugal	***	-	0,721
Romania	-		
Serbia	***	-	0,295
Slovakia	***	-	0,391
Slovenia	***	-	0,398
Spain	*	-	0,868
Sweden	-		
Switzerland	***	-	0,470
Size class (Reference: size_1 = 5 to 9 employees)			
size_2 (10 to 49 employees)	-		
size_3 (50 to 249 employees)	-		
size_4 (250 or more employees)	***	+	1,352
Sector group (Reference: Sector_4 = NACE G,H,I,R)			
NACE A	-		
NACE B, D, E, F	*	-	0,900
NACE C	-		
NACE J, K, L, M, N, S	***	+	1,425
NACE O	*	+	1,141
NACE P, Q	***	+	1,414

Table continued

Explaining factors: (NA = No answer)	Level of significance: * low (p<0,05) ** medium (p<0,01) *** high (p<0,005)	Direction of the correlation: + : more likely to discuss OSH impacts - : less likely to discuss OSH impacts	Degree/extent of the correlation (odds ratio)
Function of the respondent (Q113; Reference: Owner, managing director)			
Manager without OSH duties	-	+	1,009
Manager with OSH duties	***	+	1,168
OSH specialist	***	+	1,122
Employee representative in charge of OSH	-		
Another employee in charge of OSH	-		
External OSH consultant	-		
Ownership type (Q111; Reference: Not public)			
Public organisation	-		
NA ownership type	*	+	1,426
Type of organisation (Q100; Reference: Single-site organisation)			
Part of a multi-site organisation	***	+	1,154
OSH experts (Q151; Reference: None of the mentioned OSH expert types available in the establishment)			
Expertise	***	+	1,403
Economic situation (Q400; Reference: Very good economic situation)			
Quite good economic situation	***	-	0,832
Economic situation neither good nor bad	***	-	0,666
Quite bad economic situation	***	-	0,653
Very bad economic situation	***	-	0,499
NA economic situation	***	-	0,735
Employee representation in terms of OSH (Q350_1 to 4; Reference: non-existence of the respective position/body of representation)			
General ER (works council or trade union)			
Health and safety representation (OSH representative or co	***	+	2,064
Risk profile (Reference: 0 risks of the respective type; numerical; odds ratio per risk)			
Risk_tard (Q200_1 to _9)	***	+	1,031
Risk_psycho (Q201_1 to _7)	***	+	1,040
Visit from labour inspectorate (Q154; Reference: not visited in last 3 years)			
Visited	***	+	1,135
NA Labour inspectorate visit	-		
Dependent variable: Possible impacts of new technologies on health and safety of employees discussed? (Q311 = "yes")			

Table 22: Model 5: Dependent variable Q300: Action plan to prevent work related stress

Explaining factors: (NA = No answer)	Level of significance: * low (p<0,05) ** medium (p<0,01) *** high (p<0,005)	Direction of the correlation: + : more likely to have an action plan - : less likely to have an action plan	Degree/extent of the correlation (odds ratio)
Interview Mode (Reference: CATI)			
CAWI	-		
Country (Reference: United Kingdom)			
Austria	***	-	0,226
Belgium	***	-	0,411
Bulgaria	***	-	0,116
Croatia	***	-	0,063
Cyprus	***	-	0,140
Czech Republic	***	-	0,041
Denmark	***	-	0,709
Estonia	***	-	0,087
Finland	***	-	0,420
France	***	-	0,343
Germany	***	-	0,130
Greece	***	-	0,098
Hungary	***	-	0,121
Iceland	***	-	0,246
Ireland	***	-	0,632
Italy	***	-	0,415
Latvia	***	-	0,183
Lithuania	***	-	0,152
Luxembourg	***	-	0,104
Macedonia	***	-	0,152
Malta	***	-	0,308
Netherlands	***	-	0,284
Norway	***	-	0,280
Poland	***	-	0,099
Portugal	***	-	0,119
Romania	***	-	0,318
Serbia	***	-	0,068
Slovakia	***	-	0,078
Slovenia	***	-	0,196
Spain	***	-	0,270
Sweden	-		
Switzerland	***	-	0,233
Size class (Reference: size_1 = 5 to 9 employees)			
size_2 (10 to 49 employees)	***	+	1,134
size_3 (50 to 249 employees)	***	+	1,929
size_4 (250 or more employees)	-		
Sector group (Reference: Sector_4 = NACE G,H,I,R)			
NACE A	-		
NACE B, D, E, F	-		
NACE C	*	-	0,886
NACE J, K, L, M, N, S	***	+	1,170
NACE O	-		
NACE P, Q	***	+	1,699

Table continued

Explaining factors: (NA = No answer)	Level of significance: * low (p<0,05) ** medium (p<0,01) *** high (p<0,005)	Direction of the correlation: + : more likely to have an action plan - : less likely to have an action plan	Degree/extent of the correlation (odds ratio)
Function of the respondent (Q113; Reference: Owner, managing director)			
Manager without OSH duties	-		
Manager with OSH duties	***	+	1,262
OSH specialist	***	+	1,274
Employee representative in charge of OSH	-		
Another employee in charge of OSH	-		
External OSH consultant	*	+	1,666
Ownership type (Q111; Reference: Not public)			
Public organisation	*	-	0,900
NA ownership type	-		
Type of organisation (Q100; Reference: Single-site organisation)			
Part of a multi-site organisation	***	+	1,300
OSH experts (Q151; Reference: None of the mentioned OSH expert types available in the establishment)			
OSH expertise available	***	+	1,806
Economic situation (Q400; Reference: Very good economic situation)			
Quite good economic situation	***	-	0,874
Economic situation neither good nor bad	***	-	0,733
Quite bad economic situation	***	-	0,628
Very bad economic situation	***	-	0,608
NA economic situation	-		
Employee representation in terms of OSH (Q350_1 to 4; Reference: non-existence of the respective position/body of representation)			
General ER (works council or trade union)	***	+	1,447
Health and safety representation (OSH representative or co	***	+	1,587
Risk profile (Reference: 0 risks of the respective type; numerical; odds ratio per risk)			
Risk_tard (Q200_1 to _9)	**	-	0,978
Risk_psych (Q201_1 to _7)	***	-	0,962
Visit from labour inspectorate (Q154; Reference: not visited in last 3 years)			
Visited	***	+	1,351
NA Labour inspectorate visit	***	+	1,354
Dependent variable: Existence of an action plan to prevent work-related stress (Q300 = "yes")			

Table 23: Model 6: Dependent variable Q159: Development of sickness absence

Explaining factors: (NA = No answer)	Level of significance: * low (p<0,05) ** medium (p<0,01) *** high (p<0,005)	Direction of the correlation: + : more likely to report increased sickness absence - : less likely to report that	Degree/extent of the correlation (odds ratio)
Interview Mode (Reference: CATI)			
CAWI	-		
Country (Reference: United Kingdom)			
Austria	-		
Belgium	***	+	1,976
Bulgaria	*	+	1,279
Croatia	***	-	0,656
Cyprus	-		
Czech Republic	-		
Denmark	***	+	1,509
Estonia	-		
Finland	***	+	1,748
France	***	+	1,701
Germany	***	+	1,413
Greece	-		
Hungary	-		
Iceland	***	+	1,482
Ireland	-		
Italy	-		
Latvia	-		
Lithuania	**	-	0,715
Luxembourg	***	+	2,290
Macedonia	***	+	1,424
Malta	-		
Netherlands	***	+	2,920
Norway	***	+	2,236
Poland	-		
Portugal	-		
Romania	***	+	1,829
Serbia	-		
Slovakia	-		
Slovenia	-		
Spain	***	+	1,368
Sweden	***	+	2,059
Switzerland	***	+	1,567
Size class (Reference: size_1 = 5 to 9 employees)			
size_2 (10 to 49 employees)	**	+	1,083
size_3 (50 to 249 employees)	***	+	1,177
size_4 (250 or more employees)	***	+	1,162
Sector group (Reference: Sector_4 = NACE G,H,I,R)			
NACE A	-		
NACE B, D, E, F	-		
NACE C	*	+	1,087
NACE J, K, L, M, N, S	*	-	0,923
NACE O	-		
NACE P, Q	***	+	1,155

Table continued

Explaining factors: (NA = No answer)	Level of significance: * low (p<0,05) ** medium (p<0,01) *** high (p<0,005)	Direction of the correlation: + : more likely to report increased sickness absence - : less likely to report that	Degree/extent of the correlation (odds ratio)
Function of the respondent (Q113; Reference: Owner, managing director)			
Manager without OSH duties	-		
Manager with OSH duties	-		
OSH specialist	***	-	0,861
Employee representative in charge of OSH	***	-	0,818
Another employee in charge of OSH	***	-	0,828
External OSH consultant	-		
Ownership type (Q111; Reference: Not public)			
Public organisation	-		
NA ownership type	-		
Type of organisation (Q100; Reference: Single-site organisation)			
Part of a multi-site organisation	**	+	1,060
OSH experts (Q151; Reference: None of the mentioned OSH expert types available in the establishment)			
OSH expertise available	***	+	1,128
Economic situation (Q400; Reference: Very good economic situation)			
Quite good economic situation	*	-	0,939
Economic situation neither good nor bad	-		
Quite bad economic situation	***	+	1,295
Very bad economic situation	***	+	1,537
NA economic situation	***	-	0,729
Employee representation in terms of OSH (Q350_1 to 4; Reference: non-existence of the respective position/body of representation)			
General ER (works council or trade union)			
Health and safety representation (OSH representative or co	***	+	2,064
Risk profile (Reference: 0 risks of the respective type; numerical; odds ratio per risk)			
Risk_tard (Q200_1 to _9)	***	+	1,021
Risk_pscho (Q201_1 to _7)	***	+	1,092
Visit from labour inspectorate (Q154; Reference: not visited in last 3 years)			
Visited	***	+	1,164
NA Labour inspectorate visit	***	-	0,706
Dependent variable: Mode = Sickness absence rather increased over the last 3 years (Q159)			

Table 9: Model 7: Dependent variable Q202: Provision of ergonomic equipment (in last 3 years)

Explaining factors: (NA = No answer)	Level of significance: * low (p<0,05) ** medium (p<0,01) *** high (p<0,005)	Direction of the correlation: + : more likely to have provided ergonomic equipment - : less likely to have	Degree/extent of the correlation (odds ratio)
Interview Mode (Reference: CATI)			
CAWI	-		
Country (Reference: United Kingdom)			
Austria	***	+	1,830
Belgium	***	+	2,201
Bulgaria	***	-	0,640
Croatia	-		
Cyprus	***	+	1,325
Czech Republic	***	-	0,671
Denmark	***	+	1,723
Estonia	***	+	2,664
Finland	***	+	1,882
France	***	+	2,571
Germany	***	+	1,683
Greece	***	+	1,617
Hungary	-		
Iceland	***	+	2,784
Ireland	-		
Italy	*	+	1,157
Latvia	*	+	1,282
Lithuania	***	-	0,622
Luxembourg	**	+	1,330
Macedonia	*	-	0,785
Malta	***	+	1,592
Netherlands	***	+	2,242
Norway	***	+	1,922
Poland	***	+	1,692
Portugal	-		
Romania	-		
Serbia	-		
Slovakia	***	-	0,559
Slovenia	-		
Spain	***	+	2,214
Sweden	***	+	3,596
Switzerland	***	+	1,493
Size class (Reference: size_1 = 5 to 9 employees)			
size_2 (10 to 49 employees)	***	+	1,238
size_3 (50 to 249 employees)	***	+	1,945
size_4 (250 or more employees)	***	+	3,137
Sector group (Reference: Sector_4 = NACE G,H,I,R)			
NACE A	***	-	0,721
NACE B, D, E, F	***	-	0,848
NACE C	***	+	1,138
NACE J, K, L, M, N, S	***	+	2,529
NACE O	***	+	1,876
NACE P, Q	***	+	1,501

Table continued

Explaining factors: (NA = No answer)	Level of significance: * low (p<0,05) ** medium (p<0,01) *** high (p<0,005)	Direction of the correlation: + : more likely to have provided ergonomic equipment - : less likely to have done so	Degree/extent of the correlation (odds ratio)
Function of the respondent (Q113; Reference: Owner, managing director)			
Manager without OSH duties	***	+	1,281
Manager with OSH duties	***	+	1,533
OSH specialist	***	+	1,676
Employee representative in charge of OSH	***	+	1,411
Another employee in charge of OSH	***	+	1,350
External OSH consultant	-		
Ownership type (Q111; Reference: Not public)			
Public organisation	***	-	0,835
NA ownership type	-		
Type of organisation (Q100; Reference: Single-site organisation)			
Part of a multi-site organisation	*	+	1,061
OSH experts (Q151; Reference: None of the mentioned OSH expert types available in the establishment)			
OSH expertise available	***	+	1,713
Economic situation (Q400; Reference: Very good economic situation)			
Quite good economic situation	***	-	0,830
Economic situation neither good nor bad	***	-	0,623
Quite bad economic situation	***	-	0,472
Very bad economic situation	***	-	0,435
NA economic situation	***	-	0,691
Employee representation in terms of OSH (Q350_1 to 4; Reference: non-existence of the respective position/body of representation)			
General ER (Works Council or TU)	***	+	1,227
OSH ER (OSH rep. Or committee)	***	+	1,410
Risk profile (Reference: 0 risks of the respective type; numerical; odds ratio per risk)			
Risk_tard (Q200_1 to _9)	***	+	1,081
Risk_psycho (Q201_1 to _7)	-		
Visit from labour inspectorate (Q154; Reference: not visited in last 3 years)			
Visited	***	-	0,870
NA Labour inspectorate visit	-		
Dependent variable: Provision of ergonomic equipment as measure taken in the last 3 years (Q202)			

Table 24: ESENER-3 outcome codes by country, numerical

	Total	Country										
		AT	BE	BG	CH	CY	CZ	DE	DK	EE	EL	ES
1 No answer	26.165	1.481	138	648	276	323	7.036	625	511	132	451	2.454
2 Answer device	17.621	935	797	47	452	97	925	1.315	266	36	231	1.574
3 Busy	6.482	28	127	272	65	18	503	402	93	42	263	307
4 Information tone - Fax - Modem	13.633	45	632	111	16	117	1.505	303	166	114	418	1.390
5 Wrong telephone number	34.870	242	403	1.164	344	776	4.437	1.153	1.054	63	477	6.002
6 General appointment	51.851	1.025	954	447	489	421	4.726	10.213	1.331	127	1.343	2.703
7 Definitive appointment with target person	11.689	779	172	27	560	30	1.120	613	330	50	341	722
8 Refusal by target person	36.857	2.154	518	198	177	878	4.028	5.252	329	122	1.019	2.902
9 Refusal by contact person/reception (upfront refusal)	114.583	5.628	1.748	2.409	857	1.085	17.898	17.606	1.244	379	3.196	9.405
13 No establishment at this address (private household etc.)	6.813	35	101	127	3	295	531	714	85	23	97	1.219
14 Inactive establishment, terminated	4.365	82	47	140	32	198	247	1.126	53	33	63	146
17 Already questioned (double address)	1.950	10	110	21	12	118	268	28	104	4	16	141
18 Complete CATI interview	43.254	1.457	1.428	730	1.448	724	1.016	2.184	1.392	713	1.498	2.163
21 Stratification maximum reached (cell full)	3.119	83	129	32	40	22	78	42	178	105	294	107
22 complete, but unanalyzable	38	0	1	0	1	0	0	0	0	0	0	22
34 Refusal - add number to DO NOT CALL LIST	8.075	36	26	50	1	36	202	76	197	6	86	240
35 Partial interview, to be called back	1.361	8	33	85	4	14	21	17	20	0	88	76
36 Partial interview, not to call back	1.301	79	31	91	9	32	81	94	9	2	64	124
37 No appointment with target person possible during fieldwork time and period	8.037	833	109	24	1	26	836	498	213	24	79	1.331
38 Target person does not speak proposed languages	1.296	40	136	10	15	35	115	156	48	5	7	66
42 Size out of target (less than 5 employees or NA in Q105)	24.220	673	516	504	853	1.218	1.036	5.263	350	255	542	1.654
43 Refusal to give information in Q111, Q112 (sector information)	22	4	1	1	1	1	0	0	0	0	0	1
44 No single establishment with 5 or more employees (Q051=0)	172	0	0	2	0	9	5	0	0	10	23	0
45 Size of first contact out of scope but interview possible at subsidiary (screening countries)	29	0	0	0	0	1	1	0	0	1	3	0
46 Interview terminated after screening phase, not to call back	22	0	0	0	0	0	3	0	0	0	1	0
51 Online invitation with no result	11.820	125	523	195	275	216	2.531	544	494	128	70	1.018
52 Complete CAWI interview	2.166	46	78	25	54	33	536	80	121	45	3	103
53 Online interview rejected due to quality reasons	376	7	12	5	12	6	96	19	13	7	0	17
56 No adequate target person at the establishment	14.124	304	380	11	135	265	1.880	1.489	250	8	16	2.362
58 Online refusal	584	8	27	8	21	22	145	20	30	5	2	28
Total	446.895	16.147	9.177	7.384	6.153	7.016	51.806	49.832	8.881	2.439	10.691	38.277

	Country										
	FI	FR	HR	HU	IE	IS	IT	LT	LU	LV	MK
1 No answer	624	93	479	1.900	364	122	1.025	170	134	107	129
2 Answer device	136	1.070	57	354	553	38	305	21	61	2	29
3 Busy	33	111	352	322	65	18	164	44	280	5	4
4 Information tone - Fax - Modem	10	420	215	430	35	1	400	64	9	113	7
5 Wrong telephone number	325	1.213	234	3.132	208	236	1.441	191	265	82	60
6 General appointment	900	3.419	935	571	3.065	259	1.578	645	346	585	118
7 Definitive appointment with target person	187	401	250	742	724	71	1.219	120	79	244	13
8 Refusal by target person	595	950	132	870	290	263	581	127	236	160	34
9 Refusal by contact person/reception (upfront refusal)	1.177	2.967	2.957	2.663	349	318	6.289	606	891	624	85
13 No establishment at this address (private household etc.)	48	134	97	393	18	3	128	69	72	77	43
14 Inactive establishment, terminated	139	75	25	68	18	66	31	73	71	186	23
17 Already questioned (double address)	52	34	27	50	11	38	105	18	224	13	16
18 Complete CATI interview	1.460	2.211	722	1.475	1.986	739	2.231	733	604	734	741
21 Stratification maximum reached (cell full)	124	53	45	67	144	72	109	17	27	43	33
22 complete, but unanalyzable	13	0	0	0	1	0	0	0	0	0	0
34 Refusal - add number to DO NOT CALL LIST	29	290	181	846	439	70	34	9	105	9	4
35 Partial interview, to be called back	9	46	13	36	394	1	28	3	4	3	2
36 Partial interview, not to call back	7	67	18	54	31	6	57	4	2	8	8
37 No appointment with target person possible during fieldwork time and period	162	970	53	42	7	46	441	162	9	5	13
38 Target person does not speak proposed languages	28	37	2	46	4	22	25	31	33	2	2
42 Size out of target (less than 5 employees or NA in Q105)	543	616	172	273	375	218	322	112	411	340	92
43 Refusal to give information in Q111, Q112 (sector information)	0	1	0	0	0	0	1	0	0	0	0
44 No single establishment with 5 or more employees (Q051=0)	0	0	9	17	0	4	0	4	0	2	12
45 Size of first contact out of scope but interview possible at subsidiary (screening countries)	0	0	3	2	0	3	0	1	0	1	1
46 Interview terminated after screening phase, not to call back	0	0	3	2	0	0	0	2	0	0	0
51 Online invitation with no result	220	514	104	463	98	40	197	138	408	97	71
52 Complete CAWI interview	45	40	18	29	13	14	20	21	169	22	11
53 Online interview rejected due to quality reasons	9	8	1	3	1	6	2	4	40	3	4
56 No adequate target person at the establishment	108	435	10	913	92	19	23	114	151	82	6
58 Online refusal	10	9	0	28	3	0	9	4	16	8	2
Total	6.993	16.184	7.114	15.791	9.288	2.693	16.765	3.507	4.647	3.557	1.563

	Country										
	MT	NL	NO	PL	PT	RO	RS	SE	SI	SK	UK
1 No answer	114	396	192	1.688	298	2.759	236	295	365	410	190
2 Answer device	88	448	2.285	1.180	169	2.705	12	288	65	577	503
3 Busy	14	55	956	489	20	982	60	51	80	188	69
4 Information tone - Fax - Modem	37	205	67	2.749	251	184	19	118	93	3.346	43
5 Wrong telephone number	521	1.742	2.101	1.433	341	701	719	466	277	2.141	926
6 General appointment	452	1.147	2.582	1.844	1.563	1.308	233	1.895	326	1.106	3.195
7 Definitive appointment with target person	19	298	584	194	61	466	109	428	210	87	439
8 Refusal by target person	143	2.000	771	2.114	383	2.231	59	248	241	6.137	715
9 Refusal by contact person/reception (upfront refusal)	465	4.692	3.934	10.904	1.802	3.255	260	1.415	1.404	4.891	1.180
13 No establishment at this address (private household etc.)	117	182	107	1.095	83	476	96	5	47	172	121
14 Inactive establishment, terminated	149	32	184	328	58	165	31	122	25	222	107
17 Already questioned (double address)	67	31	58	34	45	35	25	69	13	103	50
18 Complete CATI interview	439	1.226	1.922	2.237	1.454	1.494	742	1.424	988	692	2.247
21 Stratification maximum reached (cell full)	11	319	25	175	121	75	129	65	23	0	332
22 complete, but unanalyzable	0	0	0	0	0	0	0	0	0	0	0
34 Refusal - add number to DO NOT CALL LIST	59	230	854	436	23	902	58	372	13	1.254	902
35 Partial interview, to be called back	13	2	10	25	24	40	1	8	5	27	301
36 Partial interview, not to call back	14	33	42	81	56	38	3	1	14	54	87
37 No appointment with target person possible during fieldwork time and period	13	277	1.209	132	116	72	84	53	4	181	12
38 Target person does not speak proposed languages	10	49	187	18	21	46	0	41	9	40	10
42 Size out of target (less than 5 employees or NA in Q105)	297	1.373	1.089	1.625	575	341	358	435	59	482	1.248
43 Refusal to give information in Q111, Q112 (sector information)	1	3	2	1	0	0	2	0	0	2	0
44 No single establishment with 5 or more employees (Q051=0)	10	0	0	0	40	8	8	0	7	2	0
45 Size of first contact out of scope but interview possible at subsidiary (screening countries)	3	0	0	0	8	0	0	0	1	0	0
46 Interview terminated after screening phase, not to call back	0	0	0	0	2	1	3	0	0	5	0
51 Online invitation with no result	93	1.179	98	214	249	108	71	364	286	559	130
52 Complete CAWI interview	14	295	29	13	39	6	9	88	79	64	4
53 Online interview rejected due to quality reasons	5	39	9	1	7	1	2	12	10	14	1
56 No adequate target person at the establishment	86	331	262	147	1.423	487	35	118	2	1.612	568
58 Online refusal	3	71	6	6	7	8	9	15	12	33	9
Total	3.257	16.655	19.565	29.163	9.239	18.894	3.373	8.396	4.658	24.401	13.389

10 Literature

AAPOR (2016): Standard Definitions. Final Disposition of Case Codes and Outcome Rates for Surveys. Revised version 2016; https://www.aapor.org/AAPOR_Main/media/publications/Standard-Definitions20169theditionfinal.pdf

Eurostat: ESS Standard for Quality Reports, Luxembourg 2009a

Eurostat: ESS Handbook for Quality Reports, Luxembourg 2009b

Harkness, J., Villar, A., Edwards, B. (2010): Translation, Adaptation, and Design, in: Harkness et al. (2010): Survey Methods in Multinational, Multiregional, and Multicultural Contexts; 117-140