



TNS Infratest
Sozialforschung

SECOND EUROPEAN SURVEY OF ENTERPRISES ON NEW AND EMERGING RISKS (ESENER-2)

**Open tender procedure No. EUOSHA/PRU/2012/P-03
OJ No. S235 of 06/12/2012, Tender No. 385732**

Report No. 8: Quality Report

For the

European Agency for Safety and Health at Work

Calle Santiago de Compostela 12
E-48003 Bilbao – Spain

submitted by

TNS Deutschland GmbH (TNS Infratest Sozialforschung)

Landsberger Str. 284
D - 80687 Muenchen
Germany

Your contacts:

Arnold Riedmann
Associate Director
t +49 89 5600 1183
f +49 89 5600 1441
arnold.riedmann@tns-infratest.com

Munich, 17 February 2015
67.17.132647

Contents

1	Foreword	4
2	Introduction to the statistical process and its outputs	5
2.1	General survey concept	5
2.2	Structure of the questionnaire	5
2.3	Data outputs from the statistical process	6
3	Relevance	7
4	Accuracy	9
4.1	Errors of non-observation	9
4.1.1	Coverage errors (frame errors)	10
4.1.2	Non-response errors I: Unit non-response	14
4.1.3	Non-response errors II: Item non-response	18
4.1.4	Sampling errors.....	21
4.2	Errors of observation (measurement errors)	27
4.2.1	Measurement errors deriving from the survey instrument	28
4.2.2	Measurement errors due to erroneous data given by respondents.....	29
4.2.3	Measurement errors related to the definition of respondents	30
4.2.4	Measurement errors due to interviewer influences	33
4.2.5	Effects of the mode of data collection.....	34
4.3	Processing errors.....	35
4.3.1	Data cleaning and coding errors	35
4.3.2	Weighting errors.....	36
5	Timeliness and punctuality	37
6	Coherence and comparability	39
6.1	Development of the survey concept	39
6.2	Conceptualisation and organisation of the survey	41
6.3	Usage of harmonized national questionnaire versions.....	41
6.4	International comparability – indications from the data	44
7	Respondent burden	46
8	Confidentiality, transparency and security	48
ANNEX	Regression models for analyses on mode and respondent effects51

List of abbreviations

Abbreviation	Full name/explanation		
b2b survey	Business to business survey, i.e. survey among organisations (be it at the company/enterprise or at the establishment/local unit level)		
CATI	Computer Assisted Telephone Interviewing		
CAWI	Computer Assisted Web Interviewing (online interviews)		
ESENER	European Survey of Enterprises on New and Emerging Risks		
EU-OSHA	European Agency for Safety and Health at Work		
LFS	Labour Force Survey		
OSH	Occupational Safety and Health (also called "Health and Safety at Work")		
SBR	Structural Business Register (register of business addresses compiled by each EU member state under coordination of EUROSTAT)		
SBS	Structural Business Statistics (statistics compiled by the national statistical offices on base of the SBR)		
TP	Target Person		
TripleC	TNS coordination centre for multi-country telephone surveys		
Country abbreviations (in alphabetical order)			
AL	Albania	IT	Italy
AT	Austria	LT	Lithuania
BE	Belgium	LU	Luxembourg
BG	Bulgaria	LV	Latvia
CH	Switzerland	ME	Montenegro
CY	Cyprus	MK	FYROM (Republic of 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	TR	Turkey
IS	Iceland	UK	United Kingdom

1 Foreword

This quality report relates to ESENER-2, the “**2nd European Survey of Enterprises on New and Emerging Risks**”, conducted in **2014** by TNS Infratest Sozialforschung, a division of TNS Deutschland GmbH, and the TNS network of field institutes **on behalf of the European Agency for Safety and Health at Work (EU-OSHA)**.

The report closely follows the Eurostat guidelines for quality reports on statistical data, as documented in the “ESS 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 present stage are omitted. This is for example the case for the aspects “Accessibility and Clarity”. The criteria accessibility and clarity as discussed in the “ESS Standard for Quality Reports” refer to the possibilities for other users (outside EU-OSHA) to access the collected data and to the supporting information made available for such external users. EU-OSHA already decided to make the ESENER-2 data-set accessible to interested researchers at a later stage, 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-2. In addition, the following reports are available:

- A **Technical Report**, documenting the sampling and other methodological issues, fieldwork organization, fieldwork outcomes and further information relevant for users of the ESENER-2 data-set.
- A **Translation Report**, documenting the process of the elaboration of national questionnaire versions in detail.
- A **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-2.
- A **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.

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 also analyses many of these 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.

2 Introduction to the statistical process and its outputs

2.1 General survey concept

ESENER-2 is a **cross-national survey instrument** developed for **direct collection of information of health and safety provisions and risk perceptions** from respondents **at the workplace** level. ESENER-2 was carried out as a **CATI survey**, with an **additional CAWI option** aimed at reducing non-response. 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 employee 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 covered **36 countries** (EU-28 plus Albania, Iceland, Macedonia, Montenegro, Norway, Serbia, Switzerland and Turkey).

The coordination and overall responsibility for the project on part of the contractor was with **TNS Infratest Sozialforschung** in Munich. Fieldwork itself was coordinated centrally by the international TripleC CATI centre of TNS located in Brussels, but all interviews were carried out locally by TNS partner institutes. Addresses for the survey were drawn from business address registers of different official or commercial sources. The drawing of the gross sample and the steering of the net sample were made with a multi-stratified random sampling procedure. The disproportionalities of this sample design were afterwards corrected by weighting.

ESENER-2 is the second survey of its kind conducted in Europe. The first survey, ESENER-1, was carried out in 2009. Though ESENER-1 covered basically the same topics as ESENER-2, no trend questions were taken over from the preceding survey. Also in terms of methodology, there are some differences between both survey waves.

2.2 Structure of the questionnaire

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

- **Section A1: Contact phase**
Introduction of the survey, identification of the right target person within the establishment, provision of motivation letters and scheduling of interviews
- **Section A2: Establishment and sector screening**
Screening of enterprise addresses on the existence of further local units (only for the 22 countries with no establishment-level address register); for Hungary, Montenegro and Turkey additionally screening on the sector of activity
- **Section B: Introductory questions**
Background data on the establishment and on the respondent (function within the establishment)
- **Section C: Available health and safety expertise and general policy on the topic**
Mapping of the existence of expert support and general measures taken for health promotion

- Section D: **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
- Section E: **Risk assessments**
Inquiry about the performance of risk assessments respectively the reasons for not performing them
- Section F: **Psychosocial risks and musculo-skeletal disorders**
Perception of different forms of psycho-social risks and measures taken to prevent these
- Section G: **Employee participation in health and safety issues**
- Section H: **Sources of support (on health and safety related issues)**
- Section I: **Final background questions on the establishment**

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 (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**, containing the answer codes for all questions and different types of weights.
- A gross sample file with information on all addresses used for the survey.
- **Tables** (establishment and employee proportionally weighted variants) with the results of all questions cross-tabulated **by country**.
- **Tables** (establishment and employee proportionally weighted variants) cross-tabulated **by size-class and by sector groups** (not interrelated)
- A **set of 36 tables with the results for each individual country**, in a differentiation by size-classes and sector groups (establishment proportional).

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 scientific users of the ESENER-1 data-set, including the researchers that were in charge of the various more in-depths secondary analyses on ESENER-1.
- The outcome of the "Qualitative post-test evaluation of ESENER" carried out in 5 countries (BG, DE, FI, IT and UK) by way of cognitive in-depth interviews, among others aiming at understanding how respondents of ESENER-1 interpreted and answered core questions of the survey.
- Wishes of the European Commission for the inclusion of additional aspects in ESENER-2.
- Information needs of different departments within EU-OSHA.

Not all of the recommendations and wishes for amendments could finally be taken up in the ESENER-2 questionnaire and survey concept. The inclusion of questions related to the assessment of the quality of external providers of health and safety services for example was finally discarded because pre-test results had raised doubts on the validity of the answers.

In addition to the feedback from external users of the ESENER-1 data, EU-OSHA had a series of discussions and debates with their stakeholders on the relevance of the planned ESENER-2 outcomes. Among the results of this process were decisions to implement measures for enhancing the relevance:

- A broadening of the survey population from establishments with 10 or more employees (ESENER-1) to establishments with 5 or more employees (ESENER-2). This measure leads to an increase of the relevance of the survey since this quantitatively important segment of the economy (for which most provisions on health and safety at work are equally relevant) is now also covered.
- The additional inclusion of the sector NACE Rev.2 A (Agriculture, forestry and fishing): This also enhances the relevance of the survey. Though in most European countries NACE A is not a particularly important sector in terms of the number of establishments or employees, it is characterized by a traditionally relatively high level of accidents and other health and safety risks and was thus decided to include for getting a full picture on the health and safety situation in the economy.

In terms of the processes initiated by EU-OSHA before the launch of the new survey, it can thus be stated that the further enhancement of the relevance of ESENER for all potential data users was an important issue.

While the broadening of the survey universe clearly leads to an enhancement of the relevance of the survey for a broader segment of the economy, the dropping of the employee representative questionnaire that had been part of ESENER-1 reduces the relevance of the survey results somewhat, particularly for users interested in the topic of employee representation in health and safety matters¹. This loss of relevance is compensated at least partly by broadening the section on employee representation within the ESENER-2 questionnaire and by initializing a supplementary qualitative study on employee participation (“Worker participation in the management of OSH”), based on ESENER-2.

Enhancing the relevance of the new questionnaire was also an aim of the cognitive pre-testing done before the launch of ESENER-2. There, respondents were given the chance to assess the relevance of the questionnaire for their establishment and to provide feedback on questions they considered as irrelevant. While this assessment was in general positive, there were a number of respondents from small establishments who 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.

¹ The decision to drop the employee representative questionnaire had been made with a view of enhancing the validity of the survey instrument – analyses on this part of the ESENER-1 questionnaire had shown problems with regard to the validity and international comparability inherent to the concept.

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-2 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. Groves (2004)², in his book on survey errors, 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 population³ (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

4.1 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 ob-

² Groves, Robert M. (2004): Survey Errors and Survey Costs. Hoboken, Ney Jersey: John Wiley & Sons.

³ Groves uses the term „sampling“ for the selection of a subset of the population, thus addressing the implication that statistics are estimated on the basis of the randomly selected sample (see below „sampling error“).

served. 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-2 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-2, 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 often also less frequently updated than commercial registers.

What the best choice in a country is 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 address sources used for ESENER-2 was of commercial nature (see Technical Report, Chapter 7.2 for an overview over the registers used for ESENER-2).

Over-coverage

In ESENER-2, 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 with the same person. This person would usually remember the previous call, particularly since it is a rather long survey 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 in the context of the project is therefore not an issue of concern.

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 0,5% of all addresses (n = 1.850 cases) that were touched for the survey. There are only 5 countries where this share amounts to 1,0% or more: Malta (3,7%), Latvia (2,8%), Romania (2,3%), Denmark (1,2%) and Belgium (1,0%). The three countries with the highest share of double addresses according to this criterion are all official address registers with company-level entries.

Misclassifications

Misclassifications caused by incorrect information about frame units are an issue that occurs more frequently in registers with addresses of establishments or enterprises than over-coverage. 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 (Q105) 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 immediately terminated.

The number of employees as provided in Q105 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 Q105 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 an 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 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.

Though the treatment of the misclassifications in terms of size are not a problem for the data quality, the high number of establishments screened out for size misclassifications (13.764 addresses a size classification in the address source) and the high number of 12.938 “size switchers” among the net interviews⁴ are surprising (see chapter 7.1.6 of the Technical Report for more details). This holds particularly for countries using establishment-level address registers since here, the unit listed in the register and the unit questioned for the survey is normally the same, also in case of

⁴ Among the 45.441 valid interviews in the net sample that had a size classification from the address source, 12.938 finally ended up in another size-class when considering the size indication from the respondent (Q105). The size switches 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.

units from multi-site organisations⁵. In countries that need to use company addresses, deviations between the indicated size-class and the size information provided in the interview are in turn to be expected in case of 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 (Q105) is explicitly meant to refer to the selected unit only.

b) Misclassifications due to a wrong sector attribution

Correct sector attributions are as crucial for the data quality of ESENER as correct size attributions. Therefore, at the beginning of the questionnaire (Q112) respondents were asked to confirm whether the sector attribution derived from the information in the address register is correct. If this information was considered as not correct, the respondent was asked to describe the main activity of the establishment in own words (Q113). After finalisation of the survey, these verbatims were translated into German and were centrally coded by a specialized coding team at TNS Infratest Sozialforschung. There were only few cases where the verbatim descriptions were either not valid (127 cases) or could not be clearly coded (70 cases). For these in total 199 interviews, the original code from the address sources was finally used.

All in all, 15,6% of respondents had considered the sector indication from the address source as wrong, with national results ranging from 4,4% wrong codes in Estonia to 24,2% in Switzerland. For about a third of the interviews where the sector description was considered as 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 1-digit level as indicated in the address source. Considering this, in the end only 7,9% 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 7.1 and 13.3.

Summing up, the procedure set up in the questionnaire for the verification and correction of the sector of activity and the good results of the posterior sector coding ensure that the sector attributions available in the data-set are highly reliable. The shifts between NACE Rev.2 1-digit sectors that occurred for 7,9% of all interviews led to some additional discrepancies between the targeted and the finally achieved sector structures, but these were compensated for in the weighting which for most countries considered the NACE Rev.2 1-digit structures.

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" respectively "local units". For the purpose of the study, the term "establishment" was defined as *"a company/organisation or part of a company/organisation situated in one geographically identified place"*. Not all establishment registers use the same logics and definitions for distinguishing "local units" or "establishments" from "companies" or "enterprises" on the one hand

⁵ There are however exceptions to this: Some address registers are covering the local units of multi-site organisations, but the size indication for all the units refer to the entire organisation, not only to the local establishment. In these cases (e.g. Spain), a higher share of size switchers is to be expected.

and from smaller units such as “departments” on the other hand. Larger enterprises have for example often split their different businesses into various legally independent units which are located at the same geographical site and not easily recognizable as firms of their own. These may in some registers be listed as local units, in others as companies (because they are legally independent) and in others not at all. This type of misclassification of a unit cannot easily be controlled for in a sample survey and it is likely to occur in some 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.

For the survey, the best available establishment-level address sources were to be 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. 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 the case, for example, for the non-coverage or under-coverage of public institutions within a sector characterized by a mixture of public, private and non-profit institutions.

This type of under-coverage is hard to identify because in most countries, no reference statistics are available that show the distribution of establishments by size, sector and ownership type. The share between public and private ownership within a sector may vary largely between countries, depending e.g. on how services such as public transport or the health and social work system are organized. Also, questions about the ownership type may be misunderstood by respondents or refusal rates may largely differ between different ownership types (e.g. if there is a general prohibition of certain public institutions to participate in any surveys).

There are however indications in the data-set that point at certain weaknesses in some of the registers used for the survey:

- The address source for Cyprus turned out to have only few addresses of publicly owned institutions in NACE P (Education) and Q (Health and social work).
- The address sources used for Malta, Spain, Albania and Italy are also relatively weak in the coverage of public entities within NACE P and Q.

- The address source(s) used for Albania are weak in the coverage of public institutions even in NACE O (Public administration).

For a further survey wave, these weaknesses inherent to the registers used for sampling in the concerned countries need to be taken into account and measures for improvement need to be sought, though this is not an easy task because addresses are usually not classified by the type of ownership in any of the available registers.

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 since in tendency, with higher overall response rates the risk of any response biases can supposed to be lower.

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

- The basis for the calculation of the **cooperation rate** is the sum of all addresses used for the survey **which are eligible** according to the information in the address register (i.e. those who actually have 5 or more employees and are in fact establishments and not private households).
- The basis for the calculation of the response rate is the sum of all addresses used for the survey, i.e. addresses that turned out to be ineligible or those with unknown eligibility (Wrong telephone number, No answer, Answering device etc.).


In a multi-country survey like ESENER-2 that has to use sampling frames with different quality levels, the cooperation rate is more important for a judgment on the quality of fieldwork than the response rate because non-responses due to wrong telephone numbers, fax numbers (instead of telephone numbers) etc. are mostly a sign of the quality of the address registers that were available for the survey and not primarily of fieldwork quality.


Table 1 below illustrates the mode of calculation for the cooperation and the response rate and additionally also for the refusal and contact rate. The latter indicates the share of addresses used for the survey which could be successfully contacted during fieldwork (regardless of whether the result of this contact was an interview, a refusal or another non-response reason).

The average cooperation rate for ESENER-2 (measured over all countries) is 22%. The rates for the individual countries vary considerably, ranging from 11% in Poland to 51% in Malta. The country differences have various reasons. One of them is the general attitude of potential respondents towards surveys conducted at the workplace, but other factors such as the importance of the topic in national debates may also play a role.

Table 1: Definition of response, cooperation and other rates

Code	Description	Cooperation rate	Response rate	Contact rate	Refusal rate
		% completed interviews from eligible contacts	% completed interviews from all touched addresses	% "successful" contacts from all touched addresses	% Refusals from all touched addresses
1	No answer				
2	Answer device				
3	Busy				
4	Information tone - Fax - Modem				
5	Wrong telephone number				
6	Callback possible (general appointment)				
7	Definitive appointment with target person				
8	Refusal by target person				
9	Refusal by contact person/reception (upfront refusal)				
13	No establishment at this address (private household etc.)				
14	Inactive establishment, terminated				
17	Already questioned (double address)				
18	Complete telephone interview				
20	System error				
21	Stratification maximum reached (cell full); only addresses with at least 1 call effort				
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				
41	Refusal also of online interview				
42	Size out of target (less than 5 employees or NA in Q105)				
44	No single establishment with 5 or more employees (Q051=0)				
45	Size of first contact out of scope but interview possible at subsidiary (screening countries)				
47	Refusal to provide information in the screening phase				
48	Online questionnaire sent (still open)				
52	Completed online interview received				
53	Incomplete online interview received				
56	No adequate target person at the establishment				

 Base for the calculation of the rate (denominator)

 Units counted in the calculation of the rate (enumerator)

Cooperation and response rates in ESENER-2 vary considerably between size-classes and sectors. Over all countries, it was in tendency more problematic to reach the very small establishments than the large ones. In terms of sectors of activity, the Public Administration (NACE O) stands out with a considerable better cooperation rate, while the Real estate (NACE L) and the Construction sector (NACE F) have the lowest rates (see Chapter 8.5 of the Technical Report for more details about cooperation and response rates).

Table 2: Cooperation, response, refusal and contact rates, by country

Country	Cooperation rate		Response rate		Contact rate		Refusal rate	
	Total universe (5+ employees)	Establishments with 10+ empl.	Total universe (5+ employees)	Establishments with 10+ empl.	Total universe (5+ employees)	Establishments with 10+ empl.	Total universe (5+ employees)	Establishments with 10+ empl.
AL	38%	50%	25%	32%	73%	67%	23%	25%
AT	22%	25%	18%	22%	94%	95%	33%	32%
BE	35%	37%	23%	25%	76%	77%	19%	18%
BG	30%	33%	18%	23%	70%	73%	22%	20%
CH	21%	26%	15%	22%	94%	94%	36%	33%
CY	21%	26%	10%	19%	74%	82%	20%	25%
CZ	16%	18%	10%	12%	73%	74%	31%	30%
DE	13%	16%	10%	13%	87%	87%	49%	50%
DK	37%	40%	27%	31%	83%	84%	7%	5%
EE	37%	39%	30%	34%	88%	88%	11%	11%
EL	32%	34%	24%	29%	81%	90%	8%	8%
ES	21%	23%	12%	13%	74%	72%	28%	26%
FI	32%	35%	28%	32%	97%	98%	20%	16%
FR	26%	27%	20%	22%	84%	85%	27%	26%
HR	26%	28%	22%	25%	91%	91%	18%	15%
HU	12%	14%	8%	9%	76%	76%	36%	35%
IE	18%	21%	15%	18%	93%	94%	47%	44%
IS	35%	42%	26%	32%	87%	87%	26%	22%
IT	24%	28%	16%	21%	75%	78%	34%	31%
LT	30%	32%	26%	28%	92%	92%	23%	21%
LU	28%	27%	22%	22%	91%	91%	22%	20%
LV	31%	32%	23%	25%	85%	87%	10%	9%
ME	15%	27%	6%	11%	49%	44%	22%	17%
MK	42%	51%	24%	30%	69%	69%	2%	2%
MT	51%	52%	36%	44%	97%	97%	18%	20%
NL	22%	27%	17%	21%	86%	83%	34%	30%
NO	23%	25%	15%	17%	72%	74%	32%	31%
PL	11%	18%	7%	12%	75%	78%	37%	34%
PT	38%	41%	32%	36%	91%	95%	14%	12%
RO	18%	23%	10%	13%	73%	70%	22%	25%
RS	29%	35%	22%	27%	81%	82%	20%	16%
SE	27%	30%	21%	23%	89%	89%	27%	26%
SI	28%	30%	26%	29%	98%	99%	18%	12%
SK	22%	24%	12%	14%	58%	61%	26%	24%
TR	14%	no size info	4%	no size info	40%	no size info	21%	no size info
UK	24%	25%	19%	22%	96%	96%	15%	6%
ALL	22%	26%	14%	19%	75%	82%	27%	26%

All rates shown in this report were defined in accordance with the AAPOR guidelines, though with two important exceptions:

- For both the cooperation and the response rate, we assumed that all addresses selected for the survey would be eligible for the interview. This assumption is based on the address information (sector and size) from the address provider. As shown previously, this address information is not always correct and sometimes it is even not available. A considerable number of addresses finally turned out to have less than 5 employees and to thus actually not be eligible for the survey. AAPOR proposes to estimate a rate of eligibility for these cases, but the rules on how to do these estimates are not very clearly defined for surveys among organisations⁶.

⁶ On one of the difficulties of calculating estimates on the eligibility is that the ineligibility dropouts are defined at different stages: Whereas Code 13 (private household) can usually be attributed immediately after a contact with the called number has been established and is thus also available for those who might have refused participation anyway, the ineligibility of addresses due to Code 42 (size out of target) or Code 44 (No single establishment with 5 or more employees) can be defined only within the questionnaire, i.e. after the person had already agreed to be interviewed.

- Response code 21 lists addresses that were touched and are in principle eligible, but were not used any more at a certain point because the corresponding cell of the sampling matrix had filled up in the meantime so that the address was not needed any more (stratification maximum reached). The decision to “abandon” these addresses and classify them to code 21 was partly based on the size/sector information provided in the address, partly an interview was actually started and it turned out in Q105 that the address in fact belonged to a cell of the sampling matrix that was already full.

In principle, the addresses that were not needed any more can be classified as ineligible. But at the end of fieldwork, the system shifts also addresses from soft appointments (code 6) and some other non-response codes to this category so that within this category, the ineligible addresses can hardly be distinguished from the eligible ones.

Due to these two aspects, cooperation rates resulting from the calculations are in tendency too low because all touched addresses with Code 21 were part of the denominator in the calculation of the rates.

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-2, 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

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

Within the sampling and weighting design chosen for ESENER-2, 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 7 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 intended number of interviews could be reached in spite of the selective non-response.

Selective non-response between single NACE Rev.2 1-digit sectors that were summarized to a sector group in the sampling matrix is also not a major problem for the survey since for the weighting, in most countries a finer differentiation by the 19 NACE Rev.2 sectors and the 4 size-classes was applied. In the weighing process the achieved number of interviews in a cell is being put into ratio with the real number of establishments in that cell (universe). 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 enough interviews are available for the concerned sectors. For ESENER-2, selective non-response of this type can for example be observed in the sector group NACE B, D, E, F where establishments of the Construction sector

(NACE F) participated less often than those of the other 3 groups in the sector (see Technical Report, chapter 8.5.4 for a detailed analysis of non-response by size and sector).

For unequal non-response within the cells of the weighting matrix, the situation is however 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 allowed by their headquarters to take part in any survey and are 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 one of the subsidiaries.

Ad 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-2 that asks organisations about issues that are partly legal obligations or “good practice” for a good employer. 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. Also, the reassurance of strict confidentiality and anonymous data analysis 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 (e.g. on the share of doing adequate 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.

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 indicates that the question was frequently not well understood or that the respondent understood the question, but cannot or does not want to provide an answer.

The share of item non-response in ESENER-2 is generally low: Calculated per interview, the item non-response rate in the data-set is 1,4%. This means that on average, a respondent did not answer 1,4% of the questions⁷ he or she received.

Table 3 below shows that for the single questions, item non-response rates are ranging from a very low value of 0,1% in Q104, Q107 and Q164a to 8,0% in Q115, the question asking for the year of foundation of the establishment. For the latter, a high item non-response had been anticipated and people answering with "Don't know" in Q115 were additionally Asked Q115x which repeated the question, but offered categories instead of asking for an open numerical answer. After offering categories, the overall item non-response for Q115/Q115x dropped to 0,5%.

Further questions with a comparatively high share of "No answer" (3% or more) are:

Q156:	Existence of a specific OSH budget	3,9%
Q165:	Visit from the Labour Inspectorate in last 3 years	4,7%
Q254:	Year of the last workplace risk assessment carried out	4,0%
Q256_3:	Sharing of risk assessment results with works council	3,9%
Q245_4:	Sharing of risk assessment results with trade union rep.	4,9%
Q261_3:	Risk assessment not carried out because too burdensome	4,8%
Q305:	Involvement of employees in measures on psychosocial risks	3,3%
Q307:	State of information regarding the inclusion of PSR in RA	4,8%
Q400_1:	Usage of OSH information from employers organisations	5,2%
Q400_2:	Usage of OSH information from trade unions	3,0%
Q400_3:	Usage of OSH information from insurance providers	3,3%
Q400_6:	Usage of OSH information from other official OSH institutes	3,5%
Q450:	Level of absenteeism compared to other establishments	3,8%
Q451:	Rating of the current economic situation of the establishment	3,4%

The enhanced item non-response rates in these questions are partly owed to the heterogeneity of the definition of respondents for the survey: While a manager answering the survey may e.g. not be very clear about the sources of information used by the health and safety specialist in the establishment, an employee representative in charge of OSH answering the survey may not be sure about the sources used (additionally) by the employer side.

For the CAWI interviews, the item non-response rate is higher: Taking all CAWI interviews into consideration (i.e. although the 303 CAWI interviews not accepted for the data-set), the item non-response share for CAWI is 5,6%. This value is however influenced by the high item non-response value of 19,3% in the rejected interviews. Among the 1.289 accepted CAWI interviews, the average item non-response is at 2,5%. For CAWI, this is a very low item non-response rate, suggesting that the questionnaire was all in all easy to answer.

⁷ In questions with item batteries where each item had to be answered separately, each item was counted as one question for this analysis.

Table 3: 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 in %
Q100	49.320	410 (410/0)	0,8%
Q101	34.715	196 (196/0)	0,6%
Q102	27.025	212 (162/50)	0,8%
Q103a	11.195	133 (133/0)	1,2%
Q103b	4.093	41 (41/0)	1,0%
Q104	49.320	51 (51/0)	0,1%
Q107	49.320	42 (42/0)	0,1%
Q110	49.320	1.084 (1.084/0)	2,2%
Q111	49.320	191 (191/0)	0,4%
Q112	49.320	102 (102/0)	0,2%
Q113	7.181	28 (28/0)	0,4%
Q114	49.320	210 (210/0)	0,4%
Q115	39.279	3.152 (108/3.044)	8,0%
Q115x	3.044	187 (187/0)	6,1%
Q150_1	49.320	233 (233/0)	0,5%
Q150_2	49.320	565 (565/0)	1,1%
Q150_3	49.320	643 (643/0)	1,3%
Q150_4	49.320	614 (614/0)	1,2%
Q150_5	49.320	652 (652/0)	1,3%
Q155	49.320	333 (333/0)	0,7%
Q156	49.320	1.929 (1.929/0)	3,9%
Q157	49.320	227 (227/0)	0,5%
Q158_1	49.320	225 (225/0)	0,5%
Q158_2	49.320	213 (213/0)	0,4%
Q158_3	49.320	167 (167/0)	0,3%
Q158_4	49.320	205 (205/0)	0,4%
Q160	49.320	1.254 (1.254/0)	2,5%
Q161	15.435	368 (368/0)	2,4%
Q162	25.884	243 (243/0)	0,9%
Q163	25.884	422 (422/0)	1,6%
Q164a	32.331	46 (46/0)	0,1%
Q164b	16.814	34 (34/0)	0,2%
Q165	49.320	2.333 (2.333/0)	4,7%
Q166_1	46.298	774 (774/0)	1,7%
Q166_2	44.804	429 (429/0)	1,0%
Q166_3	49.320	324 (324/0)	0,7%
Q166_4	46.769	463 (463/0)	1,0%
Q200_1	49.320	189 (189/0)	0,4%
Q200_2	49.320	83 (83/0)	0,2%
Q200_3	49.320	92 (92/0)	0,2%
Q200_4	49.320	217 (217/0)	0,4%
Q200_5	49.320	177 (177/0)	0,4%
Q200_6	49.320	75 (75/0)	0,2%
Q200_7	49.320	106 (106/0)	0,2%
Q200_8	49.320	85 (85/0)	0,2%
Q200_9	49.320	110 (110/0)	0,2%
Q201_1	49.320	266 (266/0)	0,5%
Q201_2	49.320	327 (327/0)	0,7%
Q201_3	49.320	804 (804/0)	1,6%
Q201_4	49.320	546 (546/0)	1,1%
Q201_5	49.320	255 (255/0)	0,5%
Q201_6	49.320	112 (112/0)	0,2%
Q201_7	49.320	94 (94/0)	0,2%
Q202	47.438	661 (661/0)	1,4%
Q250	49.320	485 (485/0)	1,0%
Q251	38.981	109 (109/0)	0,3%
Q252_1	38.981	542 (542/0)	1,4%
Q252_2	17.303	109 (109/0)	0,6%
Q252_3	38.981	526 (526/0)	1,3%
Q252_4	38.981	497 (497/0)	1,3%
Q252_5	38.981	1.009 (1.009/0)	2,6%
Q252_6	38.981	609 (609/0)	1,6%
Q253a	5.863	150 (150/0)	2,6%
Q253b	10.810	79 (79/0)	0,7%
Q254	38.981	1.551 (207/1.344)	4,0%
Q255	38.774	931 (931/0)	2,4%
Q256_1	38.774	585 (585/0)	1,5%
Q256_2	26.211	523 (523/0)	2,0%
Q256_3	13.451	522 (522/0)	3,9%

Question	No. of answers (unweighted)	No. of item non-response (NA/DK*)	Share of non-response in %
Q256_4	12.205	592 (592/0)	4,9%
Q256_5	38.774	1.037 (1.037/0)	2,7%
Q258b	38.981	462 (462/0)	1,2%
Q259	38.981	468 (468/0)	1,2%
Q261_1	9.854	212 (212/0)	2,2%
Q261_2	9.854	86 (86/0)	0,9%
Q261_3	9.854	473 (473/0)	4,8%
Q261_4	9.854	258 (258/0)	2,6%
Q262	9.854	171 (171/0)	1,7%
Q263_1	3.094	31 (31/0)	1,0%
Q263_2	3.094	31 (31/0)	1,0%
Q263_3	3.094	37 (37/0)	1,2%
Q264_1	49.320	419 (419/0)	0,8%
Q264_2	49.320	503 (503/0)	1,0%
Q264_4	49.320	730 (730/0)	1,5%
Q264_5	49.320	278 (278/0)	0,6%
Q264_6	49.320	452 (452/0)	0,9%
Q265_1	49.320	340 (340/0)	0,7%
Q265_2	49.320	446 (446/0)	0,9%
Q265_3	49.320	296 (296/0)	0,6%
Q265_4	49.320	342 (342/0)	0,7%
Q265_5	49.320	396 (396/0)	0,8%
Q265_6	49.320	360 (360/0)	0,7%
Q265_7	49.320	792 (792/0)	1,6%
Q300	27.307	691 (691/0)	2,5%
Q301	27.307	479 (479/0)	1,8%
Q302	15.832	266 (266/0)	1,7%
Q303_1	49.320	1.057 (1.057/0)	2,1%
Q303_2	49.320	1.179 (1.179/0)	2,4%
Q303_3	49.320	1.101 (1.101/0)	2,2%
Q303_4	49.320	1.190 (1.190/0)	2,4%
Q304	31.942	544 (544/0)	1,7%
Q305	31.942	1.054 (1.054/0)	3,3%
Q306a_3	38.592	995 (995/0)	2,6%
Q306a_4	38.592	731 (731/0)	1,9%
Q306a_5	38.592	719 (719/0)	1,9%
Q306a_6	38.592	830 (830/0)	2,2%
Q307	49.320	2.344 (2.344/0)	4,8%
Q308_1	24.600	96 (96/0)	0,4%
Q308_2	26.679	303 (303/0)	1,1%
Q308_3	49.320	635 (635/0)	1,3%
Q308_4	49.320	355 (355/0)	0,7%
Q350	34.944	277 (277/0)	0,8%
Q351	33.025	246 (246/0)	0,7%
Q352_1	11.938	104 (104/0)	0,9%
Q352_2	11.938	126 (126/0)	1,1%
Q352_3	11.938	56 (56/0)	0,5%
Q352_4	11.938	219 (219/0)	1,8%
Q352_5	11.938	160 (160/0)	1,3%
Q354	29.827	431 (431/0)	1,4%
Q356_1	49.320	493 (493/0)	1,0%
Q356_2	19.680	136 (136/0)	0,7%
Q356_3	49.320	776 (776/0)	1,6%
Q356_4	24.600	131 (131/0)	0,5%
Q356_5	49.320	312 (312/0)	0,6%
Q357	4.685	53 (53/0)	1,1%
Q358	49.320	754 (754/0)	1,5%
Q400_1	49.320	2.570 (2.570/0)	5,2%
Q400_2	49.320	1.495 (1.495/0)	3,0%
Q400_3	49.320	1.649 (1.649/0)	3,3%
Q400_5	49.320	1.396 (1.396/0)	2,8%
Q400_6	49.320	1.735 (1.735/0)	3,5%
Q401	49.320	223 (223/0)	0,5%
Q450	49.320	1.895 (1.895/0)	3,8%
Q451	49.320	1.679 (1.679/0)	3,4%
Q452	21.580	447 (447/0)	2,1%
Q453	49.320	475 (475/0)	1,0%
Q454	49.320	0 (0/0)	0,0%

* As separate category, DK (Don't know) was offered only in Q115 and Q254

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 chosen set of 36 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 which 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-2 can be characterized as a multi-stratified random sampling procedure. In this, the universe is being divided into several strata. In the case of ESENER-2, these are defined by size and sector (4 size classes and 7 sector groups, in total 28 strata). 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 proportionality⁸.
- Within the size-classes, the sector targets were then set in an establishment proportional way. Exceptions to this are only Slovenia and the United Kingdom where some sector disproportionalities were introduced for the additionally ordered boost samples.

Within each stratum, samples were drawn at random from the selected address registers, with a ratio of 5 addresses per net interview for the initial sample release. If a local fieldwork team ran out of addresses for particular cells, then new addresses were provided as replacement selectively 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 universe. For the analysis of the situation in larger size-classes, the empirical basis would be very weak. In the 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 very large establishment-proportional weighting factors in the small size-classes on which the universe of establishments is concentrated. 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

⁸ The target for each of the four size classes was defined by its percentage share of the total universe of establishments plus the percentage share of the total universe of employees, divided by 2 (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 $(40\%+20\%)/2 = 30\%$).

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 22 of the 36 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 two units (usually the headquarters and one of the subsidiaries) for each multi-site enterprise, while in countries using establishment-based registers far more than two units may be selected for interview within an organization that has numerous sites. Thus, subsidiary units of multi-site enterprises are generally 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 by way of the weighting: establishments 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'. 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 36 economies covered by the survey differ largely in size: In Malta and Montenegro, the universe of establishments 5+ to be covered by the survey consists of just about 4.000 establishments. In Germany, the largest EU economy, it amounts to more than a million and is about 290 times larger than the universe in Malta and Montenegro.

With an estimated total of almost 6.7 million establishments and 183 million employees, the ESENER-2 universe is large. Data on the total universe of establishments and employees (including establishments with 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-2 was of 49.100 interviews. In each single country, the targeted sample size was reached and in a number of countries it was even slightly surpassed so that in total, 49.320 interviews are available. With this, ESENER-2 can be considered as a large-scale survey. There are not many European-wide country surveys among organisations that have larger samples than that.

The overall sampling rate in ESENER-2 was 0,007. Or, in another perspective: On average, every interview in ESENER-2 represents ca. 135 establishments.

The national sample sizes were to a certain degree adapted to the size of the respective country and varied between 450 interviews in the smallest countries (MT and ME) to 2.250 in the largest economies. For Slovenia, Spain and the United Kingdom, 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 average⁹ represents 506 establishments while one interview in Iceland represents just 8 establishments. An interview done in Germany thus on the average represents 63,25 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. In the sizing of the national samples, therefore always a tradeoff has to be made.
- 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.

⁹ 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 4: Sampling rates, by country

Country	Establishments with 5 or more employees in NACE Rev.2 sections A-S	Net sample	Sampling rate (Net interviews/Universe)	Average factor (No. of establishments represented by 1 interview (on average))
AL	14.000	750	0,054	19
AT	134.000	1.503	0,011	89
BE	111.000	1.504	0,014	74
BG	75.000	750	0,010	100
CH	170.000	1.511	0,009	113
CY	12.000	751	0,063	16
CZ	103.000	1.508	0,015	68
DE	1.144.000	2.261	0,002	506
DK	87.000	1.508	0,017	58
EE	19.000	750	0,039	25
EL	85.000	1.503	0,018	57
ES	454.000	3.162	0,007	144
FI	68.000	1.511	0,022	45
FR	657.000	2.256	0,003	291
HR	33.000	751	0,023	44
HU	103.000	1.514	0,015	68
IE	51.000	750	0,015	68
IS	6.000	757	0,126	8
IT	674.000	2.254	0,003	299
LT	42.000	774	0,018	54
LU	12.000	752	0,063	16
LV	25.000	753	0,030	33
ME	4.000	452	0,113	9
MK	14.000	750	0,054	19
MT	4.000	452	0,113	9
NL	178.000	1.519	0,009	117
NO	70.000	1.513	0,022	46
PL	344.000	2.257	0,007	152
PT	143.000	1.513	0,011	95
RO	125.000	756	0,006	165
RS	38.000	752	0,020	51
SE	140.000	1.521	0,011	92
SI	20.000	1.051	0,053	19
SK	62.000	750	0,012	83
TR	611.000	2.251	0,004	271
UK	839.000	4.250	0,005	197
Total	6.671.000	49.320	0,007	135

The structure of the net sample

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

- Several 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 O). This is partly related to weaknesses in the sample sources (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: When targets were defined, many national fieldwork partners had requested a lowering of the targets for the largest size-class(es) because of the limited universe available in this size-class and due expected difficulties to get interviews there. In the course of fieldwork it turned out that large establishments were often more willing to participate in the survey than expected. Therefore, in the course of fieldwork targets for the largest size-class were partly raised again and those for smaller sizes accordingly lowered. Over all countries, the originally agreed targets for 250+ were met and even slightly surpassed, though not in each single country.
- As regards the sectors, general observations can hardly be made: Though some single countries had particular difficulties with sector A, O, P or Q, this cannot be generalized - in other countries, the sectors NACE O, P and Q ran particularly well and had above average response rates.

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 49.320 interviews, the confidence interval is very small for all statements based on unfiltered questions. Its maximum – not considering design effects – is about +/- 0,4, meaning that the true value of a finding lies in a span defined by the measured percentage

value +/- 0,4%. The total number of $n = 49.320$ observations is thus large enough for quite precise statements on issues calculated over the whole sample (e.g. % of all 49.320 establishments doing regular risk assessments).

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

- The 36 national net samples consist of 450 to 4.250 interviews.
- In the 19 sectors (NACE Rev.2 1-digit level), between 241 (NACE B=Mining and quarrying) and 8.722 interviews (NACE C = Manufacturing) were made.
- In the 4 size-classes, between 4.808 (250+ employees) and 20.829 (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.

Table 5 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.

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% probability¹⁰ 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 $n=2.250$ interviews¹¹.

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.

¹⁰ 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.

¹¹ 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).

Table 5: 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
n=100	80,0%	4,0%	7,9%	72,1%	87,9%
n=450	80,0%	1,9%	3,7%	76,3%	83,7%
n=750	80,0%	1,5%	2,9%	77,1%	82,9%
n=1.000	80,0%	1,3%	2,5%	77,5%	82,5%
n=1.500	80,0%	1,0%	2,0%	78,0%	82,0%
n=2.250	80,0%	0,8%	1,7%	78,3%	81,7%
n=5.000	80,0%	0,6%	1,1%	78,9%	81,1%
n=20.000	80,0%	0,3%	0,6%	79,4%	80,6%
n=49.000	80,0%	0,2%	0,4%	79,6%	80,4%
n=100	50,0%	5,0%	9,8%	40,2%	59,8%
n=450	50,0%	2,4%	4,6%	45,4%	54,6%
n=750	50,0%	1,8%	3,6%	46,4%	53,6%
n=1.000	50,0%	1,6%	3,1%	46,9%	53,1%
n=1.500	50,0%	1,3%	2,5%	47,5%	52,5%
n=2.250	50,0%	1,1%	2,1%	47,9%	52,1%
n=5.000	50,0%	0,7%	1,4%	48,6%	51,4%
n=20.000	50,0%	0,4%	0,7%	49,3%	50,7%
n=49.000	50,0%	0,2%	0,4%	49,6%	50,4%

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-2 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.

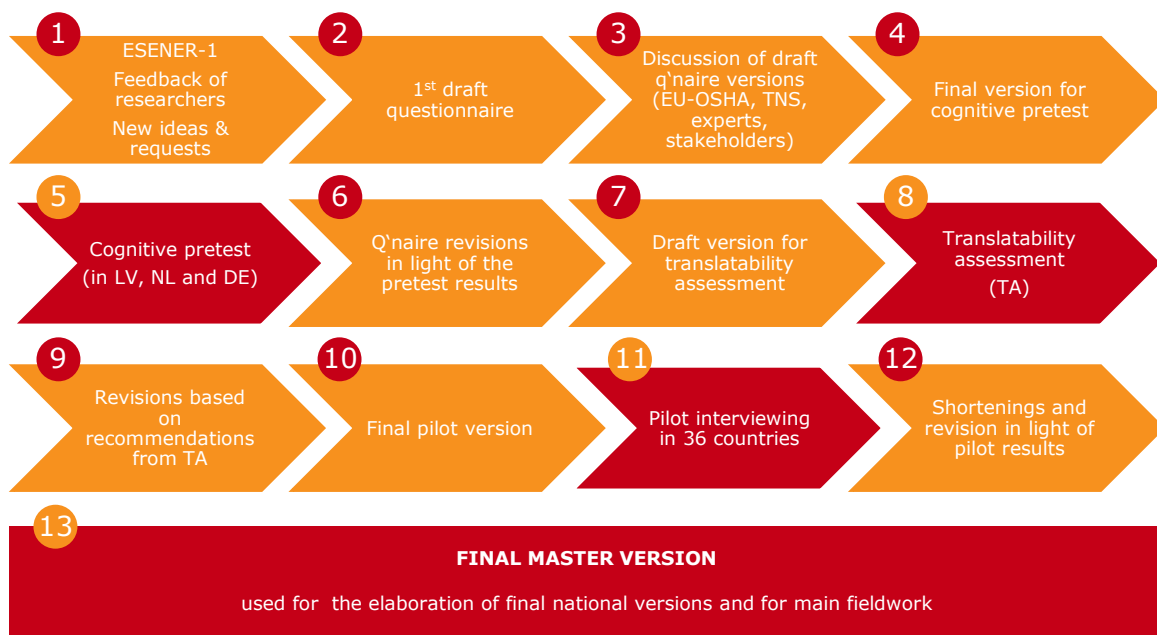
4.2.1 Measurement errors deriving from the survey instrument

In the development of the survey instrument, the EU-OSHA team, TNS Infratest and experts from different countries (Latvia, Netherlands, Germany) and institutions (Kooperationsstelle IFE GmbH Hamburg, TNO Netherlands, 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 40 establishments in Latvia, the Netherlands and Germany. 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 4 languages from different language groups, aiming at the identification of questions or formulations that may cause difficulties for translation, e.g. due to ambiguities.
- (3) A pilot survey conducted in all 36 countries, with 50 or 70 interviews per country (70 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: Main steps 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 which respondents might not be able to answer spontaneously. Additionally, in order not to urge respondents to give an answer if they were 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 are provided to EU-OSHA in a separate documentation.

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 that characterise the employer as a ‘good’ or ‘bad’ employer or are even related to the implementation of legal obligations. In spite of the assurance about the anonymity of their answers given at the very beginning of the survey, in ESENER-2 such socially desirable answers may occasionally have occurred, particularly for questions which are 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:

Q157	Regular medical examinations arranged for employees (in countries where this is obligatory)
Q166_3/_4	Formal employee representation in OSH issues (in countries/size-classes where this is obligatory)
Q250/Q255	Risk assessments and their documentation
Q265	Barriers for addressing health and safety issues
Q300-302	Action plans/procedures to deal with different types of psychosocial risks
Q356	Information 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. Unfortunately, ESENER-2 was not accompanied by an additional questionnaire variant being asked to employees or to employee representatives as some kind of verification of the answers provided by the people responsible for health and safety at the establishment level. The social desirability effect may vary between countries and between the different types of respondents of ESENER-2, depending e.g. on the stability of employment relations or the general trust in the confidentiality of answers provided in a survey.

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-2 data-set. In these, it was checked whether the function of the respondent in the establishment (Q100) 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. It can for example be observed that compared to owners, site managers or other managers not considering themselves as health and safety experts, the interviewed employee representatives in charge of health and safety were by about 14% less likely to indicate that regular risk assessments or employee trainings on the adjustment of the workplace are carried out. It might be that the employee representatives are more critical with regard to what they consider as risk assessment or a training measure taken in their establishment.

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. In the preparatory phase of the survey there were intense debates about the most appropriate definition of the respondent. For ESENER-2, finally *"the person who knows best about health and safety in this establishment"* was defined as preferred target person. In order to help with the identification of the proper respondents, additional hints on who this person could be were provided in the questionnaire text, 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. The information in the questionnaire and in the survey manual helped however to get through to the right persons. The rather low item non-response rates measured for ESENER-2 are an indicator 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 Q100_6 "External OSH consultant" (see also Q006, a question inserted to persuade respondents to answer the questionnaire locally instead of referring it to their external OSH consultant). 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 6 shows, all in all n = 175 interviews were carried out with such external OSH consultants. This is 0,4% of the total sample and thus a very small group. Relatively high absolute numbers of interviews with this group of respondents were made in Italy (36), Spain (21) the Czech Republic (18), Romania (17) and Germany (14). In relative terms (as share of the total national sample), Romania, Italy and the Czech Republic 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 2,2% (Romania) 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.

Table 6: Number and share of interviews done with external OSH consultants, by country

Country	Number of interviews with external OSH consultants (Q100_7 = 1 respectively Q100 = 6)	Share of interviews with external OSH consultants (Q100_7 = 1 respectively Q100 = 6)
AL	0	0,0%
AT	1	0,1%
BE	3	0,2%
BG	1	0,1%
CH	1	0,1%
CY	0	0,0%
CZ	18	1,2%
DE	14	0,6%
DK	0	0,0%
EE	1	0,1%
EL	1	0,1%
ES	21	0,7%
FI	0	0,0%
FR	2	0,1%
HR	0	0,0%
HU	6	0,4%
IE	0	0,0%
IS	0	0,0%
IT	36	1,6%

Country	Number of interviews with external OSH consultants (Q100_7 = 1 respectively Q100 = 6)	Share of interviews with external OSH consultants (Q100_7 = 1 respectively Q100 = 6)
LT	0	0,0%
LU	1	0,1%
LV	4	0,5%
ME	1	0,2%
MK	0	0,0%
MT	1	0,2%
NL	2	0,1%
NO	0	0,0%
PL	12	0,5%
PT	3	0,2%
RO	17	2,2%
RS	10	1,3%
SE	5	0,3%
SI	4	0,4%
SK	4	0,5%
TR	2	0,1%
UK	4	0,1%
ALL:	175	0,4%

Other than expected, the interviews with external consultants do not concentrate on small establishments for which the out-sourcing of the entire OSH duties to external suppliers is considered as a more common phenomenon than for large establishments.

Table 7: Number and share of interviews done with external OSH consultants, by size

Country	Number of interviews with external OSH consultants (Q100_7 = 1 respectively Q100 = 6)	Share of interviews with external OSH consultants (Q100_7 = 1 respectively Q100 = 6)
5 to 9 employees	20	0,2%
10 to 49 employees	67	0,3%
50 to 249 employees	58	0,5%
250 or more employees	30	0,6%
ALL	175	0,4%

Influence of the respondent type on the answers to the questionnaire

The most important question for an assessment of the new respondent definition chosen for ESENER-2 is whether the broader variety regarding the OSH background of respondents as 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-2 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:	Q356_1	Employee training on proper use & adjustment of their equipment
Model 3:	Q356_3	Employee training on psychosocial risks
Model 4:	Q259 ¹²	Risk assessment considered as useful procedure (opinion question)
Model 5:	Q300	Existence of an action plan on work related stress

While models 1, 2, 3 and 5 are analysing the respondent behaviour with regard to factual questions, model 4 examines a question which is considered to contain a good part of personal assessment or opinion on part of the respondent. Models 1, 2 and 3 are models calculated for the entire net sample. For model 5, only interviews from establishments with 20 or more employees were considered because the question was filtered on this criterion. Model 4 was calculated only for establishments that indicated to regularly carry out risk assessments (Q250 = 1). For each model, both a standard model and a stepwise model were calculated. The step-wise model shows the "ranking" of the 12 independent variables as explanatory factors for each of the analysed dependent variables.

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 (Q100, as single-punch)
- Single-site vs. multi-site organisation ("more_est" = part of a multi-site organisation)
- Usage of OSH expertise (if any of Q150_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 q16634)
- Number of traditional risks existing at the workplace (number of risks Q200_1 to _9 with "yes")
- Number of traditional risks existing at the workplace (number of risks Q201_1 to _7 with "yes")¹³

¹² The question was: "In your establishment, is the risk assessment procedure seen as a useful way of managing health and safety?"

- Visited by the labour inspectorate in last 3 years

The results of these regression analyses are documented in detail in the Annex. Summarizing, the influence of the type of respondent on the answers to the questions selected as dependent variables in the 5 models can be described as follows:

- In model 1 (regular risk assessments), there are several highly 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 as well as internal or external OSH experts are more likely to indicate the performance of risk assessments; for the specific managers and the internal OSH experts this correlation is highly significant, though with an odds ratio of 1,33 respectively 1,35 not very strong; for the external experts, the correlation is less significant, but with an odds ratio of 3,13 it is very strong.
- In models 2 and 3 (provision of training to employees), the type of the respondent has only a small influence on the answers (with employee representatives being less likely to indicate training on the workplace set-up and external consultants being more likely to indicate that training on the avoidance of psycho-social risks is performed).
- In model 4 (usefulness of risk assessments), managers without operative OSH duties as well as OSH experts are more likely to assess risk assessments as a useful procedure than owners, managing directors or other types of general workplace managers; but this correlation is significant only at the 5% level and with an odds ratio of 1,19 it is not very strong
- In model 5 (action plan on work-related stress), again managers without 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"; again, this correlation is significant only at the 5% level and is not very strong

4.2.4 Measurement errors due to interviewer influences

The survey was carried out by TNS institutes or by partner institutes of the TNS network. In each country, the survey was conducted centrally from a CATI studio by trained interviewers. 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 measures: A webex training held before the pilot survey and a 2-day training seminar held in Munich.

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 live into at least 10% of interviews and contact attempts. By this way, any undesired in-

¹³ Deviating from this general rule, for Model 5 not the number of psycho-social risks was used as independent variable, but instead only the existence of time pressure as the factor supposed to be most closely related to "work-related stress" and thus to the set-up of an action plan on this.

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-2 team.

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, 797 interviewers were working on the project. On the average, thus $49.320/797 = 62$ interviews were done per interviewer. The number of different interviewers appointed to the project was in each country small enough to develop routine and expertise in this specific questionnaire, but large enough as to ensure that any individual interviewer effect that might persist in spite of the careful selection and supervision does not have an overly influence on the overall data.

To sum, 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, in the multi-variate regression models described in the previous chapter 4.2.3 and annexed to this report also a variable for the mode of data collection was included among the independent variables.

The results of the regression models show that:

- Within the chosen model, the mode does not show any statistically significant effect on the indication of whether or not regular risk assessments are carried out in the establishment.
- The mode has also no effect on the respondent's answer to the question whether or not an action plan on work-related stress exists in the establishment.
- With regard to the models 2 and 3 related to the OSH training for employees, some mode influences can be observed: Respondents from the CAWI interviews are less likely to indicate that such training measures for employees are actually taken in the establishment. While this influence is of only low significance and not very strong in Model 2 (training on the proper use and adjustment of the equipment), it is highly significant and with an odds ratio of 0,723 also stronger for the training measures related to the prevention of psycho-social risks at work. The odds ratio of 0,723 can be interpreted as a by 28% lower probability for CAWI respondents to state the existence of this type of training measures as compared to a CATI respondent.
- In the calculated Model 4, there is no mode effect visible as far as the assessment of risk assessments as a useful instrument is concerned¹⁴.

¹⁴ In model 4 as annexed to this report, only the answers 1 (yes, considered as useful) and 2 (no, not considered as useful) were taken into account. Answers with code 3 (there are conflicting views about this within the establishment) and 9 (No answer) were excluded from the analysis. If repeating the analysis in a modified way, with answer options 2 (not useful) and 3 (conflicting views on usefulness) being summarized and opposed to answer option 1 (useful), the result of the analysis changes: Here, CAWI respondents proved to be much less likely to answer positively, i.e. to consider risk assessments as a useful procedure. The influence is however mostly not attributable to social desirability effects, but results from the fact that in the CATI version option 3 (there are conflicting views) was not to be read out actively while in the CAWI questionnaire it was visible to all respondents.

4.3 Processing errors

4.3.1 Data cleaning and coding errors

In computer assisted surveys, posterior data cleaning and editing is hardly necessary. Various data checks carried out with the CATI and CAWI scripts before the launch of fieldwork and weekly data checks during fieldwork, with a programmed syntax, ensured that filtering mistakes were minimized.

The only technical mistake that occurred at the beginning of the survey was not an error in the programmed script, but was related to a mistake in the preparation of the net sample for Poland: Due to a wrong code in the address file, in Poland question Q102 (single-vs. multi-site) was not asked for the first 110 cases.

Three further issues related to the data are worth mentioning here, though these are not related to processing errors:

- In **Estonia**, no term for "Works Council" was asked in Q166. Instead, in Q166_1, instead of the term for "Works Council", the term for the "Health and Safety Committee" was asked. The "Health and Safety Committee" should have been asked in Q166_4, but this item was left empty for Estonia on advice of national experts. In the data processing after finalisation of the survey, the term for "Health and Safety Committee" was then set from Q166_1 to Q166_4 which is the place where the "Health and Safety Committee" was asked about in all other countries. For the item on the "Works Council", unfortunately no data are available. This item (Q166_1) was therefore left empty for Estonia.
- In **Greece**, national experts had advised not to ask for the existence of "Works Councils" since this body was considered as hardly relevant for firms in Greece. After the first interviews, the Greek fieldwork institute did however recommend to ask this term also for Greece. The term was therefore amended in the Greek questionnaire and establishments that had been interviewed already with the original questionnaire version were called back in order to ask for the existence of a "Works Council". The answers were added to the data file so that there is no restriction regarding the interpretation of the Greek data on Q166.
- In **Bulgaria**, due to a mistake that occurred in the translation process (and also slipped through in the later questionnaire checks done by national health and safety experts), only one of two terms proposed as appropriate Bulgarian equivalents of a "Works Council" were asked. The term that was asked is very broad (translating rather as "assembly of workers") and there was a surprisingly high share of affirmative answers to this question. Due to doubts about the validity of the answers, all data on the concerned question (Q166_1) and on relevant follow-up questions using it as a filter were set to sysmis (missing data) for Bulgaria.

The ex-post coding of the sector descriptions provided by respondents who considered the sector attribution from the address as wrong (Q112/Q113 in the questionnaire) was done by a small team of professionals 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 (e.g. 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. Details of the weighting, including measures on the effectiveness of the weighting, are set out in the Technical Report (chapter 11) 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 to be based on assumptions. These assumptions 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-2 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 – partly – Romania).

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. Nevertheless, the quality of universe structures for which estimates were necessary is still leading to more accurate results than the usage of company statistics for a survey for which the statistical unit and the unit of enquiry were defined as "establishments/local units".

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 14th June 2013, 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. In turn, other steps were finalized earlier than scheduled.

The most important target date was the delivery of the finalized, integrated and weighted data-set scheduled for 28 November 2014. This date was exactly met, with no delay.

Table 8: Time periods for key project steps (plan, reality and comments)

Key working steps	Comments	Planned start	Planned finalisation	Actual finalisation
2013				
Inception meeting	The inception meeting marks the start of the project work	14.6	14.6	14.6
Development of a first draft questionnaire version	Contrary to the original planning, a first draft questionnaire was not yet provided by EU-OSHA, but was drafted after the project start in cooperation between EU-OSHA and TNS Infratest.	24.6	1.7	2.7
Finalisation of the cognitive test instruments	The finalisation of the cognitive test instruments (questionnaire including cognitive questions; translation of the questionnaire into Latvian, German and Dutch; development of test guidelines and reporting templates)	5.9	9.9	10.9
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.	16.9	4.10	11.10
Revision of questionnaire and finalisation of master version for translatability check	The revision needs after the cognitive pre-test were less than anticipated so that the process of the further questionnaire revision could be done quicker than anticipated.	21.10	8.11	23.10
Translatability assessment	The translatability assessment could be started earlier than planned and was carried out in a shorter time period.	11.11	22.11	7.11
Revision of the questionnaire in light of the findings from the translatability assessment	Started and terminated earlier than originally planned.	25.11	6.12	18.11

2014				
Preparation of translation tools and training of translators and verifiers	Slightly later finalisation of the last of the series of training sessions (one per language group plus reserve session); some translators (trained in an earlier session) had also started translating in the meantime.	9.12	10.1	13.1
Finalisation of translation and verification; verified national versions sent to EU-OSHA for checks by domain experts	The translations took slightly longer than planned so that the time buffer from the earlier finalisation of previous step was needed	2.1	17.2	14.2
Finalisation of translation process, including integration of feedback from domain experts and final proofreading	Due to some delays in the collection of feedback from the domain experts (late comers) and the need for clarifications in part of the feedback, this step took about one week longer than anticipated.	14.2	21.3	28.3
Programming and testing of the international master script and all national script versions	Programming of the master script started in parallel with the translation process; after finalisation of the translations, the CATI master script was over-written with national language versions. Process largely as scheduled.	2.1	25.4	29.4
Preparation (training, set-up of script and monitoring instruments etc.) and fieldwork for the pilot survey	2,5 days delay in the start of the pilot fieldwork because the adaptation of the script to the TripleC CATI environment (server system, sample management system, address take-up for 2nd interviews in screening countries) took a few days longer than anticipated. Fieldwork shortened by 2-3 days in order to keep up time schedule. Targeted number of pilot interviews reached nevertheless.	28.4	16.5	16.5
Final changes to the survey instrument and to the accompanying material	Discussion and implementation of changes following the pre-test, both in master questionnaire and in national versions; process as scheduled	9.6	27.6	27.6
Fieldwork for main survey	The period originally foreseen for fieldwork was from 28/07/2014 to 03/10/2014. Due to the large sample boosts for ES and the UK, a prolongation of this period by 4 weeks was agreed (2 weeks before the regular start of FW in the other countries plus two weeks after regular fieldwork end); in addition, for some countries with total summer holiday closures in the fieldwork period it was agreed to also grant this longer period respectively part of it. In the end, the prolongation after the scheduled regular end of the fieldwork period had to be granted to a number of additional countries since otherwise the targets in particular cells would not have been met. Almost all countries could however finalize fieldwork by the data foreseen for the boost countries (17/10/2015). The only exception is TR which took one working day longer for finalisation (20/10/2015).	14.7	17.10	20.10
Data extraction, editing and coding; weighting of the data-set; elaboration of cross tabulations	Process largely as scheduled; but some delays in the finalisation of the weighting matrices lead to delays in the preparation of the further reporting	20.10	27.11	27.11
Delivery of weighted and integrated data-set, first set of cross-tabulations and a first rough draft of the Technical Report	Delivery of the data-set at the agreed date.	28.11	28.11	28.11
Amendments to the Technical Report and delivery of a final report version	Delivery as scheduled.	1.12.	19.12	19.12
2015				
Elaboration and delivery of further technical reports and documentation on the survey	Delivery of final reports and documentation for January agreed in November/December.	1.12	30.1	1.2

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, TNS Infratest 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-1 conducted in 2009 in 31 countries.

For ESENER-2, 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 ESENER-2, a decision on the establishment/local unit level had already been taken by EU-OSHA before the launch of the project. 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 14 of the 36 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 22 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. It 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.

Usage of international classifications

The usage of international classifications is an important prerequisite to establish international comparability of a survey. In ESENER-2, 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.

Other than in ESENER-1 which was conducted only about 1,5 years after the introduction of the new version of the NACE classification (NACE Rev. 2), there was no country that still used the old NACE Rev.1 classification in either the address register or in the statistical background information.

Some of the address sources used for the survey (e.g. the Dun & Bradstreet register used for the UK) 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-2, 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) was however somewhat higher than average for the UK and other countries where SIC codes had to be transformed into NACE.

This slight 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 address source if it was considered wrong. It had however some repercussions on the preciseness to which the intended sample structures were met because the verbatims on the sector descriptions of respondents were coded only after finalisation of the survey and the revised codes could thus not be used in the monitoring of the sample. In the preparation phase of the survey, this drawback of the ex-post coding solution was discussed and the advantages of a fully homogenous central coding by a coding expert team (and not by interviewers with little experience in this) were considered to outweigh slight deviations of the sector structures in the final samples as compared to the calculated targets.

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.

6.2 Conceptualisation and organisation of the survey

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 (with a chance for local fieldwork partners to provide feedback on their national version),
- 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 propose modifications for targets considered as 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:

- 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)
- Spot checks on the interviews and other quality control measures on the performance of interviewers (e.g. analyses of the rate of successful contacts per interviewer)
- 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 usage 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 usage 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 36 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 Q166_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 Q166 do not exist in all of the countries. In Germany and Austria, for example, the general 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 Q166 – forms of representation not relevant in a country were not asked there.

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, TNS Infratest Sozialforschung cooperated with cApStAn Linguistic Quality Control, an institute specialized in the translation and verification of questionnaires and other material for high quality cross-national surveys. The entire translation process was handled by cApStAn, with additional supervision and support by TNS Infratest Sozialforschung.

The translation process consisted of several steps that were meant to ensure a high degree of comparability between all language versions. In the following, the main steps are shortly summarized (the process is documented in more detail in the Translation Report):

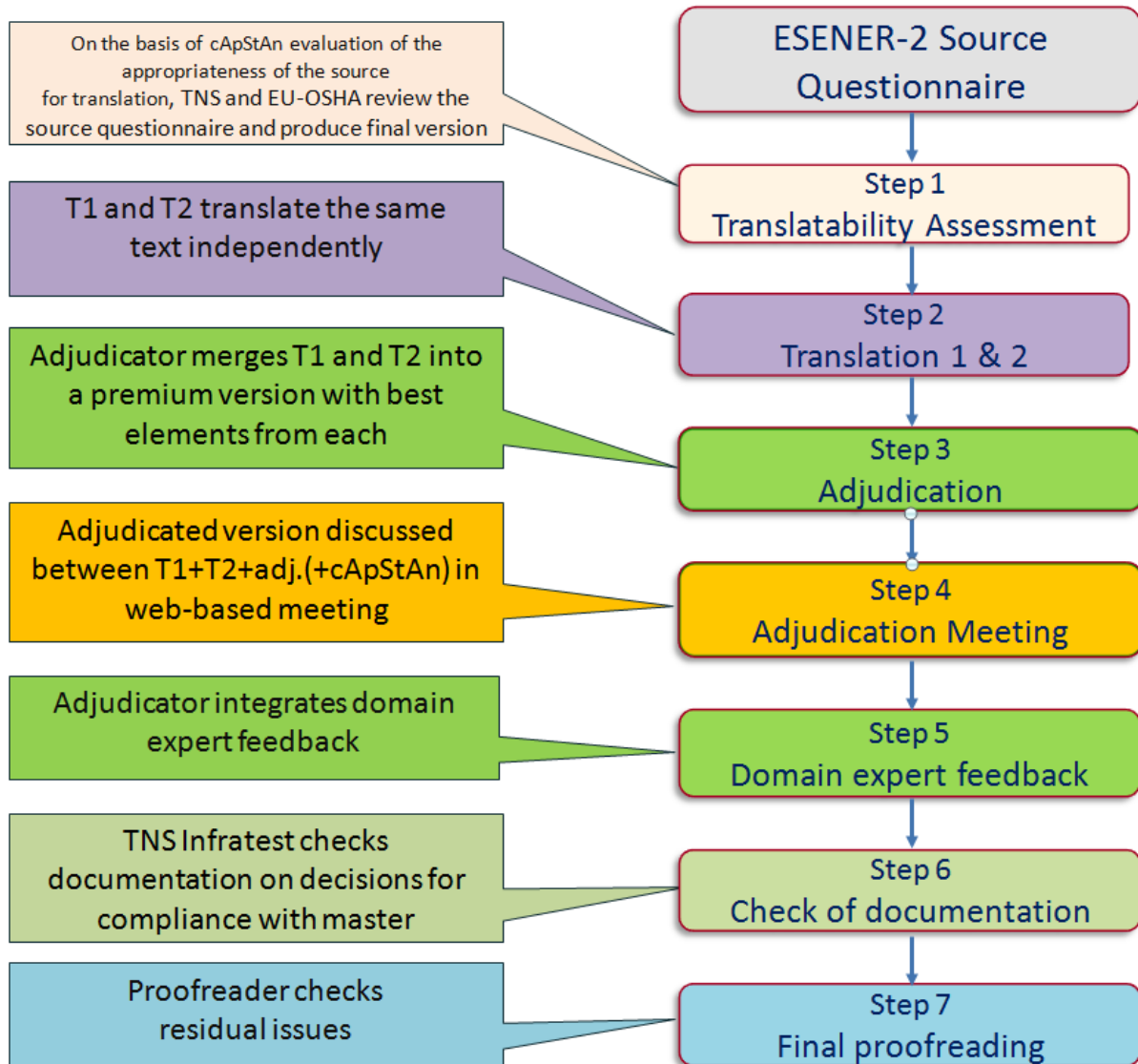
- A few weeks before the launch of the translation process, a translatability assessment was carried out in order to detect formulations in the English master questionnaire which are difficult to translate into other languages or might cause ambiguities. To this end, translators from four languages of different language families (Czech as Slavonic language, French as a Romance language, Swedish as Germanic/Nordic language and Greek) elaborated rough translations of the master version and documented any difficulties encountered hereby. The results were analysed by linguists at cApStAn and proposals for modifications on the master questionnaire were made and implemented.
- In a first step of the genuine translation process, then two independent translations were produced by different translators of the cApStAn translator pool (T1 and T2).
- The two translations were reviewed by a third translator, called adjudicator. As adjudicators, particularly experienced translators were used. 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.

- For all language versions, review meetings with the participation of the adjudicator, both translators and a moderator were held, mostly in form of webex meetings. In these review meetings, the decisions of the adjudicator were discussed. For aspects for which there was no clear and easy decision, a common solution was sought during the review meeting.
- The adjudicated versions were then sent to national health and safety experts of the Focal Point network of EU-OSHA where they were checked again, with a particular focus of the correctness of all specific health and safety terminology used in the questionnaire.
- The feedback of the Focal Point experts was reviewed by the adjudicators and implemented (provided that the proposals were in line with the intentions of the master questionnaire).
- In addition, the TNS Infratest team checked the (English language) comments on the adjudicator's decisions between translation 1 and 2 and on the proposals from the domain expert. The focus here was on whether the decisions taken were in line with the intentions of the master questionnaire.
- 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.

The outcome of this last step was 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 made¹⁵. The entire process including all decisions and changes on individual questions or items has been documented for each country.

¹⁵ There is one exception to this, though it is not related to the translations as such: In Greece, there was a change of the national language term used for "Works Council" during fieldwork, see chapter 4.3.1 for more details on this.

Figure 2: Summary of the translation process¹⁶



6.4 International comparability – indications from the data

In many countries, ESENER is the only survey among organisations on the topic of health and safety at work. Only relatively few countries (e.g. LV; DK and DE) have conducted representative own national 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 dif-

¹⁶ Graph taken from the translation report elaborated by cApStAn: Roberta, Lizzi/Krajceva, Elica/Dept, Steve: Translation and Adjudication Report. ESENER-2 (Enterprise survey on new and emerging risks); Brussels 2014, Annex 10, p.30; unpublished

ferent 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.

Some first analyses done with the ESENER-2 data do not show any totally unexpected results. Further analytical work with the data will show whether there are any indications on restrictions as regards the comparability of the results.

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 24,22 minutes for completion. This average is however calculated by summing up the average duration from all countries and dividing this by the number of countries (36), 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 20,000 hours. Assuming an 8 hour working day as standard, this is equivalent to about 2.500 person days.

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 this preparatory time, measurements are not available.

On the part of an individual participant, the burden is with about 24 minutes of average interviewing time at the limit of what is in many countries considered as acceptable for most respondents of the targeted group. 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.

Table 9: Respondent burden in terms of interviewing time, by country

Country	Mean duration (minutes)	Number of interviews	Total time burden in minutes	Total time burden in hours
al	19,18	750	14.385	239,8
at	24,19	1.503	36.359	606,0
be	26,18	1.504	39.374	656,2
bg	27,95	750	20.959	349,3
ch	28,39	1.511	42.892	714,9
cy	19,24	751	14.448	240,8
cz	26,03	1.508	39.250	654,2
de	25,86	2.261	58.475	974,6
dk	23,44	1.508	35.351	589,2
ee	23,87	750	17.906	298,4
el	23,52	1.503	35.344	589,1
es	24,79	3.162	78.395	1.306,6
fi	29,37	1.511	44.381	739,7
fr	26,44	2.256	59.648	994,1
hr	25,82	751	19.389	323,2
hu	22,22	1.514	33.643	560,7
ie	18,66	750	13.991	233,2
is	23,07	757	17.466	291,1
it	23,71	2.254	53.434	890,6
lt	23,34	774	18.069	301,1
lu	28,23	752	21.229	353,8
lv	28,20	753	21.231	353,9
me	22,01	452	9.947	165,8
mk	19,14	750	14.355	239,3
mt	30,59	452	13.827	230,5
nl	28,31	1.519	42.997	716,6
no	22,42	1.513	33.925	565,4
pl	24,35	2.257	54.969	916,1
pt	24,70	1.513	37.372	622,9
ro	23,36	756	17.660	294,3
rs	22,52	752	16.932	282,2
se	25,80	1.521	39.248	654,1
si	21,49	1.051	22.581	376,3
sk	21,06	750	15.793	263,2
tr	23,43	2.251	52.751	879,2
uk	21,13	4.250	89.814	1.496,9
ALL:	24,22	49.320	1.194.644	19.963,2

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. 6 months after fieldwork finalisation, 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 (deletion is foreseen after a period of 18 months after fieldwork finalisation). There were two questions related to the allowance of a re-contact:

Q453:

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

Q454 (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, such as hints on the right to withdraw this allowance at any time and the indication of an address for this.

Respondents were also 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 countries).

Further measures were necessary to inhibit an ex-post identification of individual firms on base of the interview data:

- The data on the size of the actual workforce (Q104 and Q105) 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 was not provided in the data-set. There, only the NACE Rev.2 1-digit level is made available. For the data-set that will be made accessible to external users outside EU-OSHA, the some small sectors NACE B, D and E will additionally be summarized to a sector group B D E in order to prevent de-anonymisation of the data.
- Likewise, data collected on the distribution of establishments by region (within a country) was not included in the data-set.

Annex:

Logit regression models for analysis of mode and respondent effects

Table 10: Modell 1: Dependent variable Q250 (Risk Assessments)

Explaining factors:	Level of significance: * low (p<0,05) ** medium (p<0,01) *** hoch (p<0,005)	Direction of the correlation: + : more likely to carry out risk assessments - : less likely to carry out RA	Degree/extent of the correlation (odds ratio)
Mode (Reference: CATI)			
CAWI_mode	-		
Size-class (Reference: size_1 = 5-9 employees)			
Size_2 (10-49 employees)	***	+	1,412
Size_3 (50-249 employees)	***	+	2,314
Size_4 (250+ employees)	***	+	3,736
Sector group (Reference: sector group 4 = NACE G,H,I,R)			
Sector_1 (NACE A)	-		1,124
Sector_2 (NACE B,D,E,F)	***	+	1,549
Sector_3 (NACE C)	***	+	1,370
Sector_5 (NACE J, K,L,M,N,S)	***	-	0,807
Sector_6 (NACE O)	***	-	0,623
Sector_7 (NACE P,Q)	-		
Ownership (Reference: Non-public organisation)			
Public	***	-	0,856
Public_na (no answer)	-		
Respondent type/function (Reference: Owner, managing director, branch manager)			
Manager_without OSH duties	*	-	0,920
Manager_with OSH duties	***	+	1,335
OSH specialist	***	+	1,355
Employee representative for OSH	***	-	0,857
External OSH consultant	**	+	3,139
Type of site (Reference: Single-site organisation)			
Multi-site organisation	***	+	1,192
OSH experts (Reference: OSH experts available in the establishment, Q150)			
Expertise	***	+	2,851
Employee representation (Reference: No employee representation of the respective type in the est.)			
General ER (Works Council or TU)	***	+	1,239
OSH ER (OSH rep. or committee)	***	+	2,391
Risk profile (Reference: 0 risks of the respective type; numerical; odds ratio per risk)			
Risk_trad (Q200_1 to _9)	***	+	1,118
Risk_psycho (Q201_1 to _7)	***	-	0,881
Visit from the labour inspectorate (Reference: not visited in last 3 years)			
Visited	***	+	1,707
Visited_na	***	-	0,737

Table continued

Country (Reference: United Kingdom)			
Albania	***	-	0,120
Austria	***	-	0,068
Belgium	***	-	0,134
Bulgaria	*	-	0,716
Croatia	***	-	0,255
Cyprus	***	-	0,069
CzechRepublic	***	-	0,220
Denmark	*	-	0,747
Estonia	***	-	0,223
Finland	***	-	0,127
France	***	-	0,150
Germany	***	-	0,104
Greece	***	-	0,089
Hungary	***	-	0,312
Iceland	***	-	0,057
Ireland	***	-	0,204
Italy	-		
Latvia	***	-	0,615
Lithuania	***	-	0,198
Luxembourg	***	-	0,038
Macedonia	***	-	0,094
Malta	***	-	0,187
Montenegro	***	-	0,149
Netherlands	***	-	0,324
Norway	***	-	0,200
Poland	***	-	0,351
Portugal	***	-	0,324
Romania	***	-	0,372
Serbia	***	-	0,238
Slovakia	***	-	0,085
Slovenia	***	+	1,721
Spain	***	-	0,648
Sweden	***	-	0,313
Switzerland	***	-	0,072
Turkey	***	-	0,370
Dependent variable: Risk assessments (regular risk assessments carried out, Q250)			

Table 11: Model 2: Dependent variable Q356_1: Training on workplace

Explaining factors:	Level of significance: * low (p<0,05) ** medium (p<0,01) *** hoch (p<0,005)	Direction of the correlation (Q356_1): + : more likely to train on proper use & adjustment of equip. - : less likely to train on that	Degree/extent of the correlation (odds ratio)
Mode (Reference: CATI)			
CAWI_mode	*	-	0,864
Size-class (Reference: size_1 = 5-9 employees)			
Size_2 (10-49 employees)	***	+	1,098
Size_3 (50-249 employees)	***	+	1,361
Size_4 (250+ employees)	***	+	1,730
Sector group (Reference: sector group 4 = NACE G,H,I,R)			
Sector_1 (NACE A)	-		0,965
Sector_2 (NACE B,D,E,F)	-		1,066
Sector_3 (NACE C)	***	+	1,112
Sector_5 (NACE J, K,L,M,N,S)	-		0,967
Sector_6 (NACE O)	-		0,954
Sector_7 (NACE P,Q)	***	-	0,841
Ownership (Reference: Non-public organisation)			
Public	***	-	0,849
Public_na (no answer)	-		0,817
Respondent type/function (Reference: Owner, managing director, branch manager)			
Manager_without OSH duties	-		0,949
Manager_with OSH duties	-		1,035
OSH specialist	-		0,982
Employee representative for OSH	***	-	0,864
External OSH consultant	-		1,067
Type of site (Reference: Single-site organisation)			
Multi-site organisation	*	+	1,058
OSH experts (Reference: OSH experts available in the establishment, Q150)			
Expertise	***	+	1,931
Employee representation (Reference: No employee representation of the respective type in the est.)			
General ER (Works Council or TU)	*	+	1,063
OSH ER (OSH rep. or committee)	***	+	1,999
Risk profile (Reference: 0 risks of the respective type; numerical; odds ratio per risk)			
Risk_trad (Q200_1 to _9)	***	+	1,087
Risk_psych (Q201_1 to _7)	***	-	0,915
Visit from the labour inspectorate (Reference: not visited in last 3 years)			
Visited	***	+	1,263
Visited_na	-		0,992

Table continued

Country (Reference: United Kingdom)			
Albania	-		
Austria	***	-	0,504
Belgium	***	-	0,225
Bulgaria	***	-	0,230
Croatia	***	-	0,384
Cyprus	***	-	0,716
Czech Republic	***	-	0,223
Denmark	***	-	0,341
Estonia	***	-	0,584
Finland	***	-	0,486
France	***	-	0,119
Germany	***	-	0,356
Greece	***	-	0,573
Hungary	***	-	0,270
Iceland	***	-	0,402
Ireland	***	-	0,653
Italy	***	-	0,658
Latvia	***	-	0,346
Lithuania	***	-	0,679
Luxembourg	***	-	0,261
Macedonia	***	-	0,461
Malta	***	-	0,470
Montenegro	***	-	0,557
Netherlands	***	-	0,258
Norway	***	-	0,642
Poland	***	-	0,138
Portugal	***	-	0,360
Romania	***	-	0,115
Serbia	***	-	0,415
Slovakia	***	-	0,554
Slovenia	***	-	0,646
Spain	-		
Sweden	***	-	0,397
Switzerland	***	-	0,450
Turkey	-		
Dependent variable: Training on the proper use and adjustment of the working equipment provided			

Table 12: Modell 3: Dependent variable Q356_3: Training on prevention of PSR (Q356_3)

Explaining factors:	Level of significance: * low (p<0,05) ** medium (p<0,01) *** hoch (p<0,005)	Direction of the correlation (Q356_3): + : more likely to train on psycho-social risks - : less likely to train on psycho-social risks	Degree/extent of the correlation (odds ratio)
Mode (Reference: CATI)			
CAWI_mode	***	-	0,723
Size-class (Reference: size_1 = 5-9 employees)			
Size_2 (10-49 employees)	-		
Size_3 (50-249 employees)	*	+	1,076
Size_4 (250+ employees)	***	+	1,541
Sector group (Reference: sector group 4 = NACE G,H,I,R)			
Sector_1 (NACE A)	***	-	0,804
Sector_2 (NACE B,D,E,F)	***	-	0,808
Sector_3 (NACE C)	***	-	0,744
Sector_5 (NACE J, K,L,M,N,S)	***	+	1,115
Sector_6 (NACE O)	**	+	1,152
Sector_7 (NACE P,Q)	***	+	2,104
Ownership (Reference: Non-public organisation)			
Public	***	+	1,161
Public_na (no answer)	-		
Respondent type/function (Reference: Owner, managing director, branch manager)			
Manager_without OSH duties	-		
Manager_with OSH duties	-		
OSH specialist	-		
Employee representative for OSH	-		
External OSH consultant	*	+	1,515
Type of site (Reference: Single-site organisation)			
Multi-site organisation	***	+	1,241
OSH experts (Reference: OSH experts available in the establishment, Q150)			
Expertise	***	+	1,706
Employee representation (Reference: No employee representation of the respective type in the est.)			
General ER (Works Council or TU)	***	+	1,257
OSH ER (OSH rep. or committee)	***	+	1,613
Risk profile (Reference: 0 risks of the respective type; numerical; odds ratio per risk)			
Risk_trad (Q200_1 to _9)	***	-	0,979
Risk_psycho (Q201_1 to _7)	-		
Visit from the labour inspectorate (Reference: not visited in last 3 years)			
Visited	***	+	1,292
Visited_na	***	+	1,210

Table continued

Country (Reference: United Kingdom)			
Albania	***	-	0,721
Austria	***	-	0,469
Belgium	***	-	0,511
Bulgaria	***	-	0,405
Croatia	***	-	0,331
Cyprus	***	-	0,709
CzechRepublic	***	-	0,280
Denmark	***	-	0,458
Estonia	***	-	0,428
Finland	***	-	0,587
France	***	-	0,426
Germany	***	-	0,319
Greece	***	-	0,733
Hungary	***	-	0,551
Iceland	***	-	0,716
Ireland	-		
Italy	-		
Latvia	***	-	0,605
Lithuania	***	-	0,624
Luxembourg	***	-	0,348
Macedonia	***	-	0,763
Malta	***	-	0,548
Montenegro	***	-	0,521
Netherlands	***	-	0,541
Norway	***	-	0,672
Poland	***	-	0,741
Portugal	***	-	0,496
Romania	***	-	0,462
Serbia	***	-	0,558
Slovakia	***	-	0,544
Slovenia	-		
Spain	-		
Sweden	***	-	0,552
Switzerland	***	-	0,452
Turkey	-		
Dependent variable: Training on the prevention of psycho-social risks at work (Q356_3)			

Table 13: Model 4: Risk assessments considered as a useful instrument

Explaining factors:	Level of significance: * low (p<0,05) ** medium (p<0,01) *** hoch (p<0,005)	Direction of the correlation (Q259): + : more likely to consider risk assessments as a useful instrument - : less likely to consider it as useful	Degree/extent of the correlation (odds ratio)
Mode (Reference: CATI)			
CAWI_mode	-		1,047
Size-class (Reference: size_1 = 5-9 employees)			
Size_2 (10-49 employees)	-		1,049
Size_3 (50-249 employees)	-		1,117
Size_4 (250+ employees)	-		0,983
Sector group (Reference: sector group 4 = NACE G,H,I,R)			
Sector_1 (NACE A)	*	-	0,729
Sector_2 (NACE B,D,E,F)	-		0,852
Sector_3 (NACE C)	-		0,997
Sector_5 (NACE J, K,L,M,N,S)	***	-	0,817
Sector_6 (NACE O)	-		1,004
Sector_7 (NACE P,Q)	-		1,120
Ownership (Reference: Non-public organisation)			
Public	-		1,064
Public_na (no answer)	-		0,901
Respondent type/function (Reference: Owner, managing director, branch manager)			
Manager_without OSH duties	*	+	1,190
Manager_with OSH duties	-		1,235
OSH specialist	*	+	1,191
Employee representative for OSH	-		1,111
External OSH consultant	-		2,032
Type of site (Reference: Single-site organisation)			
Multi-site organisation	**	+	1,170
OSH experts (Reference: OSH experts available in the establishment, Q150)			
Expertise	***	+	1,512
Employee representation (Reference: No employee representation of the respective type in the est.)			
General ER (Works Council or TU)	***	+	1,295
OSH ER (OSH rep. or committee)	***	+	1,669
Risk profile (Reference: 0 risks of the respective type; numerical; odds ratio per risk)			
Risk_trad (Q200_1 to _9)	***	+	1,088
Risk_psych (Q201_1 to _7)	***	-	0,794
Visit from the labour inspectorate (Reference: not visited in last 3 years)			
Visited	***	+	1,193
Visited_na	***	+	1,643

Table continued

Country (Reference: United Kingdom)			
Albania	***	-	0,414
Austria	***	-	0,349
Belgium	***	-	0,395
Bulgaria	***	-	0,330
Croatia	-		0,682
Cyprus	*	-	0,539
Czech Republic	***	-	0,378
Denmark	***	-	0,231
Estonia	-		0,872
Finland	***	-	0,577
France	***	-	0,538
Germany	***	-	0,275
Greece	-		1,073
Hungary	***	-	0,360
Iceland	***	-	0,249
Ireland	-		0,840
Italy	*	-	0,674
Latvia	***	-	0,366
Lithuania	***	-	0,239
Luxembourg	-		0,905
Macedonia	-		0,970
Malta	-		2,094
Montenegro	-		1,179
Netherlands	***	-	0,363
Norway	***	-	0,575
Poland	***	-	0,284
Portugal	*	+	1,611
Romania	-		0,866
Serbia	***	-	0,493
Slovakia	***	-	0,463
Slovenia	***	-	0,307
Spain	***	-	0,458
Sweden	-		0,835
Switzerland	*	-	0,671
Turkey	-		1,085
Dependent variable: Risk assessment considered as useful (Q259)			

Table 14: Model 5: Dependent variable Q300: Action plan to prevent work related stress

Explaining factors:	Level of significance: * low (p<0,05) ** medium (p<0,01) *** hoch (p<0,005)	Direction of the correlation (Q300): + : More likely to have an action plan on work-related stress - : less likely to have such an action plan	Degree/extent of the correlation (odds ratio)
Mode (Reference: CATI)			
CAWI_mode	-		
Size-class (Reference: size_1 = 5-9 employees)			
Size_2 (10-49 employees)	-		
Size_3 (50-249 employees)	-		
Size_4 (250+ employees)	-		
Sector group (Reference: sector group 4 = NACE G,H,I,R)			
Sector_1 (NACE A)	-		
Sector_2 (NACE B,D,E,F)	-		
Sector_3 (NACE C)	-		
Sector_5 (NACE J, K,L,M,N,S)	***	-	0,798
Sector_6 (NACE O)	-		
Sector_7 (NACE P,Q)	-		
Ownership (Reference: Non-public organisation)			
Public	-		
Public_na (no answer)	-		
Respondent type/function (Reference: Owner, managing director, branch manager)			
Manager_without OSH duties	*	+	1,193
Manager_with OSH duties	-		
OSH specialist	*	+	1,191
Employee representative for OSH	-		
External OSH consultant	-		
Type of site (Reference: Single-site organisation)			
Multi-site organisation	*	+	1,122
OSH experts (Reference: OSH experts available in the establishment, Q150)			
Expertise	***	+	1,488
Employee representation (Reference: No employee representation of the respective type in the est.)			
General ER (Works Council or TU)	***	+	1,273
OSH ER (OSH rep. or committee)	***	+	1,676
Risk profile (Reference: 0 risks of the respective type; numerical; odds ratio per risk)			
Risk_trad (Q200_1 to _9)	***	+	1,054
Risk_psycho (Q201_1 to _7)	***	-	0,725
Visit from the labour inspectorate (Reference: not visited in last 3 years)			
Visited	***	+	1,191
Visited_na	***	+	1,666

Table continued

Country (Reference: United Kingdom)			
Albania	***	-	0,411
Austria	***	-	0,372
Belgium	***	-	0,404
Bulgaria	***	-	0,359
Croatia	-		
Cyprus	*	-	0,535
CzechRepublic	***	-	0,396
Denmark	***	-	0,218
Estonia	-		
Finland	***	-	0,585
France	***	-	0,555
Germany	***	-	0,291
Greece	-		
Hungary	***	-	0,395
Iceland	***	-	0,263
Ireland	-		
Italy	-		
Latvia	***	-	0,372
Lithuania	***	-	0,270
Luxembourg	-		
Macedonia	-		
Malta	-		
Montenegro	-		
Netherlands	***	-	0,355
Norway	***	-	0,591
Poland	***	-	0,313
Portugal	*	+	1,596
Romania	-		
Serbia	***	-	0,534
Slovakia	***	-	0,504
Slovenia	***	-	0,314
Spain	***	-	0,488
Sweden	-		
Switzerland	-		
Turkey	-		
Dependent variable: Existence of an action plan on work-related stress (Q300)			

List of Tables:

Table 1: Definition of response, cooperation and other rates.....	15
Table 2: Cooperation, response, refusal and contact rates, by country	16
Table 3: Share of item non-response per question (unweighted results).....	20
Table 4: Sampling rates, by country	24
Table 5: Standard error (s.e.) and confidence interval (c.i.) at different levels of disaggregation ..	27
Table 6: Number and share of interviews done with external OSH consultants, by country	31
Table 7: Number and share of interviews done with external OSH consultants, by size.....	31
Table 8: Time periods for the main project phases (plan, reality and comments).....	37
Table 9: Respondent burden in terms of interviewing time, by country.....	47
Table 10: Modell 1: Dependent variable Q250 (Risk Assessments)	49
Table 11: Model 2: Dependent variable Q356_1: Training on workplace	51
Table 12: Modell 3: Dependent variable Q356_3: Training on prevention of PSR (Q356_3).....	53
Table 13: Model 4: Risk assessments considered as a useful instrument	55
Table 14: Model 5: Dependent variable Q300: Action plan to prevent work related stress.....	57