

Jointly produced by
CEDEFOP and
EUROFOUND

Labour market change
European Company Survey 2019:
Data quality assessment

[European Company Survey 2019:
Workplace practices unlocking employee potential](#)



CEDEFOP

European Centre
for the Development
of Vocational Training



Eurofound

European Foundation
for the Improvement
of Living and Working
Conditions

Authors: Sam Desiere and Karolien Lenaerts (HIVA, KU Leuven)

Research managers: Giovanni Russo (Cedefop) and Gijs van Houten (Eurofound)

Acknowledgements: This project could not have been completed without the invaluable contribution of Sophia MacGoris (Eurofound) and Christopher White (Eurofound) to the preparation and implementation of the European Company Survey 2019, which included the checking and editing of this report.

Eurofound reference number: WPEF19017

Related reports: European Company Survey 2019: Technical and fieldwork report; European Company Survey 2019: Translation report; European Company Survey 2019: Sampling and weighting report; European Company Survey 2019: Quality control report; European Company Survey 2019: Data editing report; European Company Survey 2019: Coding report

© European Centre for the Development of Vocational Training (Cedefop), 2020/European Foundation for the Improvement of Living and Working Conditions (Eurofound), 2020
Reproduction is authorised provided the source is acknowledged.

For any use or reproduction of third-party photos or other material permission must be sought directly from the copyright holders. Any queries on copyright must be addressed in writing to: copyright@eurofound.europa.eu

Research carried out prior to the UK's withdrawal from the European Union on 31 January 2020, and published subsequently, may include data relating to the 28 EU Member States. Following this date, research only takes into account the 27 EU Member States (EU28 minus the UK), unless specified otherwise.

This report presents the results of research conducted largely prior to the outbreak of COVID-19 in Europe in February 2020. For this reason, the results do not fully take account of the outbreak.

The European Centre for the Development of Vocational Training (Cedefop) is the European Union's reference centre for vocational education and training. It provides information on and analyses of vocational education and training systems, policies, research and practice. Cedefop was established in 1975 by Council Regulation (EEC) No 337/75.

The European Foundation for the Improvement of Living and Working Conditions (Eurofound) is a tripartite European Union Agency, whose role is to provide knowledge in the area of social, employment and work-related policies. Eurofound was established in 1975 by Council Regulation (EEC) No. 1365/75 to contribute to the planning and design of better living and working conditions in Europe.

European Centre for the Development of Vocational Training

Telephone: +30 2310490111

Email: info@cedefop.europa.eu;

Web: www.cedefop.europa.eu

European Foundation for the Improvement of Living and Working Conditions

Telephone: +353 1 204 31 00

Email: information@eurofound.europa.eu

Web: www.eurofound.europa.eu

Contents

Abstract.....	1
Executive summary	1
1. Introduction	4
1.1. Background to the 2019 European Company Survey.....	4
1.2. Approach to the quality assessment and structure of the report.....	5
2. Assessment of survey processes.....	8
2.1. Questionnaire development and translation.....	8
2.2. Sampling strategy, sampling frames and reference statistics.....	16
2.3. Fieldwork.....	20
2.4. Weighting	36
2.5. Data and results dissemination.....	40
2.6. The Quality Assurance Plan	42
3. Assessment of data quality.....	45
3.1. Internal validity.....	45
3.2. External validity	47
4. Recommendations.....	53
References	60
Appendix 1	62
Appendix 2	69

Abstract

The European Company Survey (ECS) is a unique, nationally representative survey on workplace practices across establishments in the EU28. The ECS has the potential to offer new insights on, among others, workplace practices, employee participation, innovation and skills (mis)match. This external quality assessment, carried out by HIVA-KU Leuven, evaluates the data quality of the ECS 2019. The assessment confirms that the survey adopted best practices. The dataset gathered from managers contains 21,869 observations and is of high quality. Compared to previous ECS waves, the main limitation of the ECS 2019 is the low number of interviews with employee representatives (3,073). This is most likely due to the switch from a phone survey to a push-to-web survey. The assessment offers recommendations to further improve the data quality of future waves. In addition, the assessment contains valuable insights for survey practitioners who consider conducting an EU-wide push-to-web survey.

Executive summary

Purpose

The European Company Survey (ECS) is a representative EU-wide survey on workplace practices in establishments operating in the private sector employing at least 10 workers. The ECS 2019, fielded from January to July 2019, gathered data from 21,869 managers (MM) and 3,073 employee representative (ER) in the 28 EU Member States. The survey was jointly carried out by Eurofound (EF) and Cedefop (CF), and was implemented by Ipsos. A key innovation of the ECS 2019 is the switch from a phone to a push-to-web survey. This report assesses the data quality of the ECS 2019 and offers recommendations for future waves.

Methodology

The assessment consists of three parts. The first part assesses the quality of the survey processes and is structured along five stages: questionnaire development, sampling frame development, fieldwork, weighting, and dissemination. This assessment relies on the detailed documentation provided by Eurofound and Cedefop including the Quality Assurance Plan (QAP), as well as the survey literature. The second part assesses the data quality of the management (MM) and employee representative (ER) datasets. It discusses the internal as well as the external validity of the ECS 2019. The final part concludes and offers recommendations.

Findings

Assessment of survey processes

The **questionnaire development and, particularly, the questionnaire translation** followed state-of-the-art procedures. The questionnaire was developed using the ECS 2013 questionnaire as starting point and was then further refined based on stakeholder and expert inputs. Using validated scales, developing detailed research plans and selecting a set of core questions that is included in each wave could further strengthen the survey. The advance translatability assessment, cognitive testing, the implementation of the TRAPD approach to translate the source questionnaire and the pilot test

Disclaimer: This working paper has not been subject to the full Cedefop/Eurofound evaluation, editorial and publication process.

all contributed to the semantic, conceptual and normative equivalence of the questionnaire across countries.

The complex **sampling strategy** was well-thought through. Two innovations improved the survey's representativeness: the inclusion of a 'no size stratum', which increased coverage in countries where the sampling frame contains many establishments of which the size is unknown; and the strategy put in place to limit remaining 'open sample' at the end of the survey. The choice of **sampling frames and reference statistics** followed best practices, but the data quality suffers to some extent from the lack of availability of establishment level sampling frames in 17 countries.

Mainstage **fieldwork** was a daunting task and suffered from two setbacks. The quality of mainstage fieldwork benefitted greatly from the insights gained from the pilot, which, for example, helped to fine-tune the contact strategy. In general, elaborate procedures were in place to avoid non-sampling errors and to monitor data quality in real time. A first setback was the lower than expected yield rates. More establishments had to be contacted than expected to reach the target number of MM interviews. Both the share of successful screener interviews (i.e. the first phone contact during which email addresses are obtained of eligible establishments) and the conversion rate (i.e. the share of respondents who completed the online survey after a successful screener) were lower than expected. The overall yield rate was 5.2% and ranged from 1.5% in Poland to 15.9% in Lithuania. Despite these low yields, the final MM sample size is 21,869, which is close to the target of 22,700. A second setback relates to the low number of completed ER questionnaires. The ER contact details were mainly collected during the screener interview, but 68% of the MM refused to share ER contact details. By the end of the survey, only 3,073 ER questionnaires of high-quality were completed. In only 1,835 establishments both the ER and the MM were interviewed. In comparison, the ECS 2013 conducted 6,739 ER interviews in the private sector. In addition to the small sample size, the ER sample also suffers from selection bias as MM are more likely to share ER contact details if they have a good working relation with the ER.

The **weighting procedures** followed standard practices. The weighting procedure adjusted for the complex survey design, non-response at the screening and the CAWI stage, and calibrated the weights so that the estimates matched key population statistics.

There is room to improve **data dissemination**. Despite that this dataset is in many ways unique, the observation that the ECS 2013 dataset was only 462 times downloaded over a period of nearly four years suggest that the data is not sufficiently exploited. New ongoing events related to the launch of the ECS 2019 may help to further disseminate the data.

Despite containing over 100 indicators, the **Quality Assurance Plan** is a useful tool to monitor survey processes and data quality throughout the survey's lifecycle. At the same time, meeting the targets is a necessary, but not a sufficient condition to obtain high-quality data.

Assessment of data quality

In general, the **MM dataset** has a high **internal validity**. The dataset can be used to compile reliable and accurate statistics at the EU-level, by geographical clusters or by welfare regimes. More care is warranted when compiling country-level statistics or when using the data to rank countries because the precision of the estimates varies by country. The **ER dataset is less reliable**. Due to the low sample size, the lack of observations in certain countries and the selection bias, the representativeness of the ER sample can be questioned, limiting its potential to compile statistics at the EU-level.

The **external validity** refers to the extent that findings from the ECS are in line with findings from other datasets measuring similar concepts. Explorative analyses show, for instance, that training increased substantially between 2013 and 2019 according to the ECS and increased only marginally according to Eurostat's indicator lifelong learning. Hence, both datasets point towards different trends. Although comparing finding of the ECS to findings based on other datasets could reveal interesting patterns, it is fair to say that the analyses in this report only explore the potential of this approach and are currently not sufficiently convincing to warrant strong claims about the data quality.

Recommendations

The assessment offers fifteen recommendations. These recommendations are structured along the survey's lifecycle (including the QAP) and are ranked according to their impact on data quality and feasibility.

We consider it essential to address two critical issues before implementing the next wave. First, a strategy should be developed to deal with the low number of ER interviews. This may require setting ER targets, switching to a mixed mode survey, no longer conducting ER interviews or conducting a separate ER and MM survey. This is a difficult decision because every option has drawbacks. Second, strategies should be put in place to increase the use of the ECS 2019 data.

Of the remaining thirteen recommendations, the following four have a medium impact and are easy to implement:

1. Determine a set of core questions that are asked in each wave so that time trends can be identified
2. Keep (most of) the questions so that the questionnaire development and translation phase can be skipped or shortened
3. Document yield rates/non-response by country, sector and firm size so that sampling design can be optimised
4. Programme an additional consistency check comparing the number of employees reported in the screener with the number reported by the MM

1. Introduction

1.1. Background to the 2019 European Company Survey

The European Company Survey (ECS) is a cross-national questionnaire-based representative sample survey among establishments employing at least 10 people on workplace practices. Data are collected from management representatives (MM) and employee representatives (ER) (if present). Because of its focus on workplace practices, the ECS is complimentary with other EU-wide surveys, such as the European Working Conditions Survey (EWCS), which assesses working conditions by surveying workers across the EU. Together, these surveys contribute to the Europe 2020 strategy of smart, sustainable and inclusive growth.

The European Company Survey was launched in 2004 as the European Establishment Survey on Working Time and Work-life Balance (ESWT), and it was repeated in 2009, 2013 and 2019. The fourth wave of the survey was jointly carried out by Eurofound (EF) and Cedefop (CF) and implemented by Ipsos. Figure 1 presents a timeline highlighting main phases and key milestones in the survey process.

Figure 1: Timeline



Source: own compilation see Annex 2

The sheer scale of the ECS is impressive (Table 1). The survey collected data in over 20,000 establishments in the 28 EU Member States on workplace practices, such as work organisation, human resource management, skills use, skills strategies, direct employee participation and social dialogue. The complexity of multi-country company surveys like the ECS requires outstanding project management. Just to give one example, throughout mainstage fieldwork 629 interviewers reached out to 419,159 establishments across the EU28 Member States to achieve 21,869 and 3,073 high-quality interviews with respectively a manager and an employee representative.

The ECS 2019 differs in several respects from the ECS 2013 and other previous waves:

1. The ECS 2019 is the first EU-wide company survey using a push-to-web methodology. The ECS 2013 was a telephone survey;
2. It is the first ECS wave that is jointly carried out by Eurofound and Cedefop;
3. It addresses several new topics, including skills strategies and skills (mis)match;
4. The sampling and weighting strategy were revised. In line with the recommendations of previous quality assessments and feasibility studies (Eurofound, 2017a), the ECS 2019 no longer includes the public sector, used more accurate and up-to-date establishment-level

sampling frames, revised the strategy to select establishments of multi-establishment sites when a company-level sampling frame was used, and, in some countries, sampled establishments of unknown size ('no size stratum') to increase coverage;

- To increase the response rate, it offered management respondents a customised report on workplace practices.

Table 1: The ECS 2019 in a nutshell

Number of countries involved	28 (+4 IPA countries* in the pilot)
Number of interviewers in mainstage fieldwork	629
Sample dialled	419,159
Email reminders sent	251,718 (MM) + 17,574 (ER)
Number of high-quality MM interviews achieved	21,869
Number of high-quality ER interviews achieved	3,073

Note: 'Instrument for Pre-Accession Assistance' countries included North Macedonia, Montenegro, Serbia and Turkey
Source: Own compilation (Eurofound and Cedefop, 2020a)

These developments make the ECS 2019 particularly interesting to assess in terms of its survey design, survey implementation and data quality. It offers lessons for future waves as well as for other (company) surveys considering switching to a push-to-web methodology.

1.2. Approach to the quality assessment and structure of the report

This report assesses the survey quality of the ECS 2019. Each stage of a survey's lifecycle is critical to the final survey quality. Following the literature (e.g. Eurostat's Generic Statistical Business Process Model and the 'Survey Lifecycle' in the *Cross-Cultural Survey Guidelines*¹), five survey stages are distinguished:

- Questionnaire development and translation
- Sampling frame development
- Fieldwork
- Weighting
- Dissemination

The first section of the quality assessment evaluates the quality of each stage of the **survey process**. To do so, we draw on (1) the Quality Assurance Plan (QAP), developed by Eurofound and Cedefop, and (2) best practices identified in the literature and in other EU-wide surveys.

A pilot study was set up prior to mainstage fieldwork. Insights from the pilot were key to the successful implementation of mainstage fieldwork. The pilot tested the questionnaire and led to improvements in the contact strategy and sampling management strategy. We do not discuss the pilot as a separate survey stage but highlight how the lessons learnt from the pilot informed each

¹ <http://ccsg.isr.umich.edu/index.php/chapters/translation-chapter/10-chapter-list/4-what-are-cross-cultural-survey-guidelines>

stage of mainstage fieldwork. Instead, we refer the reader to the ‘pilot report’ for a detailed description and assessment of the pilot.

The QAP consists of a detailed set of indicators used to monitor the survey’s quality in each stage of the survey. These indicators are linked to the five quality dimensions of the European Statistical System: (i) relevance, (ii) accuracy, (iii) timeliness and punctuality, (iv) accessibility and clarity, and (v) coherence and comparability (ESS, 2019). Following the approach taken in the quality assessment of the European Quality of Life Survey (Eurofound, 2018), we focus on the indicators related to accuracy. The QAP is structured along twelve stages², which we reorganised in five survey stages. In addition, the QAP distinguishes between required (RQ), real world (RW) and ideal world (IW) indicators. Required targets have to be achieved; real world targets can be achieved; and ideal world targets are desired but not necessarily likely to be achieved. Section 2.6 briefly discusses whether and how the QAP contributed to achieving high-data quality.

Going beyond this framework, the ECS processes are compared with best practices identified in the literature on data quality in cross-national surveys. This includes the literature on ‘Total Survey Error’ - a conceptual framework describing the statistical properties of survey statistics accounting for sampling and non-sampling errors (Groves & Lyberg, 2010) – as well as the literature emphasising the need for a ‘user perspective’ (Lyberg, 2012). In addition, literature addressing the challenges of multi-national, multi-cultural and multi-regional surveys (‘3MC surveys’) (Harkness et al., 2010) was consulted.

The second section of the quality assessment examines the **data quality of the final dataset**. We distinguish between the survey’s *internal* and *external validity*. The internal validity discusses key statistical properties such as non-response bias and survey design effects. The *external validity* examines whether findings from the ECS 2019 are in line with (1) other surveys measuring similar concepts and (2) the ECS 2013. Both checks help to gain insights into one of quality assessment’s key questions: how did the switch from a telephone survey to a push-to-web survey affect data quality?

The third section summarises **recommendations** to improve survey processes and data quality in future waves.

Eurofound and Cedefop provided us with comprehensive technical documentation as well as the final datasets. This information can be consulted on Eurofound’s website or is available on request. More specifically, among others, the following documents were consulted:

1. Tender documents: Preparation and Implementation of the 4th European Company Survey (ECS)
2. Management and employee representative questionnaire
3. Translation report
4. Pilot report
5. Cognitive interview report
6. Technical and fieldwork report
7. Coding report
8. Sampling report
9. Quality control report

² Sampling, Weighting, Questionnaire, Translation, Pilot, Fieldwork infrastructure, Scripting, Fieldwork training, Fieldwork, Data processing, Micro data, Paradata.

The assessment also relied on four feasibility studies commissioned by Eurofound prior to the launch of the 4th ECS regarding (1) the methodological lessons learnt from the 3rd ECS; (2) sampling frames and modes; (3) survey modes; and (4) the feasibility of a panel survey. The background paper for the questionnaire development provided valuable insights in skills measurement (Eurofound, 2019). The quality assessment of previous waves (Petraikos, Kleideri, & Ieromnimon, 2011; Vila & Carausu, 2013) and of the latest wave of the European Quality of Life Survey (Eurofound, 2018) were also valuable sources. The key challenges encountered throughout the survey's lifecycle were discussed with Femke De Keulenaer, Ipsos' project management lead of the ECS 2019, on March 13, 2020.

2. Assessment of survey processes

2.1. Questionnaire development and translation

The ECS is a multi-national, multi-regional and multi-cultural survey designed for comparative cross-national research. This makes the questionnaire's development and translation challenging (Harkness et al., 2010). The first step is to design a questionnaire that translates the concepts that one intends to measure using (validated) questions and scales. The second step is to translate the source questionnaire in the target languages. This translation must ensure functional equivalence; that is the questions should have the same meaning across cultures and regions (Saris & Gallhofer, 2014).

The next section discusses the development and the content of the questionnaire. We evaluate whether (validated) scales are used to measure the different concepts. We then discuss the questionnaire's translation. Based on our assessment detailed below, we conclude that the translation from English to national languages followed state-of-the-art procedures. As this process is already outstanding and well documented (Eurofound and Cedefop, 2020a), we suggest continuing the current approach. At the same time, we do see room for improvement regarding the use of validated scales.

Table 2: Accuracy indicators related to questionnaire development

No.	Indicator	Target	Results	Assessment
Questionnaire development				
3.2.(RW)	Percentage of questionnaire items in the final source questionnaire that meet international methodological standards of question design (such as outlined in Saris & Gallhofer (2007))	100%	100%	✓
Advance translation				
3.6 (RQ)	Percentage of questionnaire items where substantive ambiguities are spotted for which either the source questionnaire is adjusted or a translation instruction is drafted	100%	100%	✓
Cognitive testing				
3.9 (RW)	Number of questions for which 'major' issues were detected that were kept	0	4	✗
3.10 (RQ)	A strategy for cognitive testing is agreed that that reflects current academic standards, as illustrated by academic references included in a document outlining the strategy for cognitive testing	Yes	Yes	✓
3.13 (RQ)	Percentage of countries in which the selection of the respondents and composition of the sample corresponds with the agreed approach and design	Yes	Yes	✓
Pilot¹				
8.5 (RQ)	Percentage of translation issues detected in the pilot, for which the solution was based on the input of two independent translators	100%	100%	✓

Source: (Eurofound and Cedefop, 2020c)

Note: ¹ The pilot contains four additional QA indicators which are not directly relevant for mainstage fieldwork and which are not reported here. RQ: required indicator; RW: real world indicator

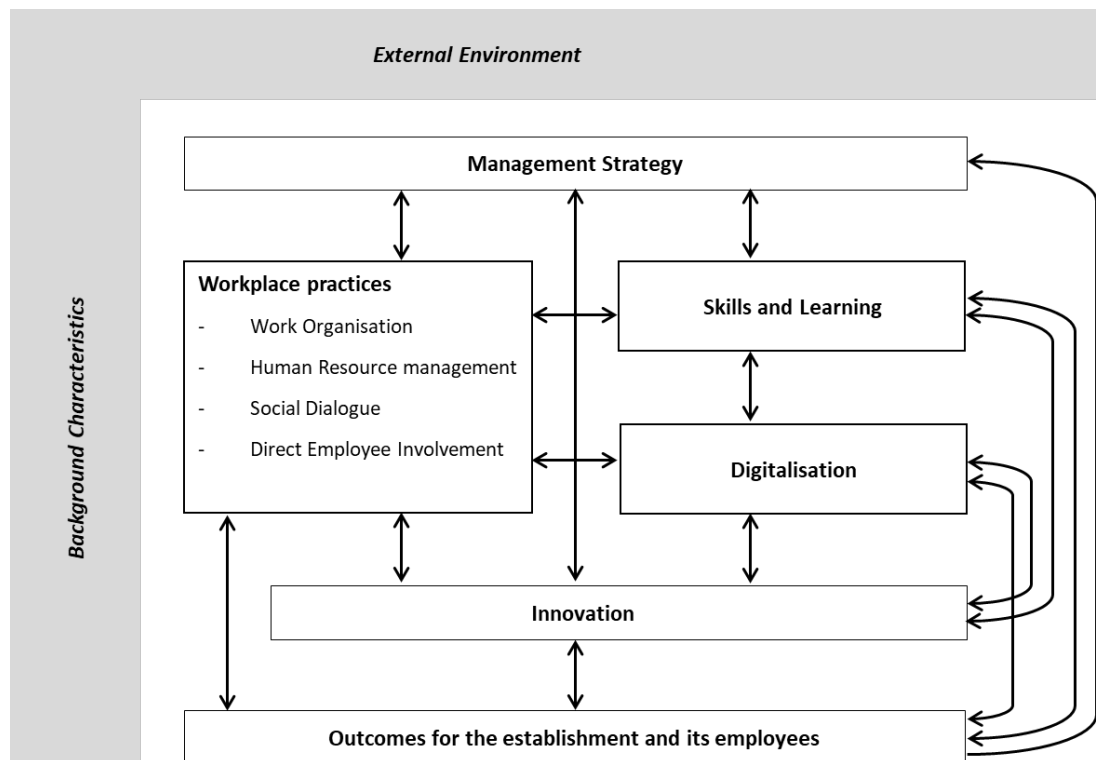
The QA indicators confirm the questionnaire's quality (Table 2). Five out of the six QA indicators related to the accuracy of the development of the questionnaire were met. The only missed target relates to four questions for which cognitive testing detected 'major' issues but that were nevertheless kept in the questionnaire. We believe that missing this target has no impact on the data quality because the problematic questions were revised. Two of the four problematic questions were replaced by questions fielded in the ECS 2013, while the two remaining questions were rephrased (Eurofound and Cedefop, 2020b, p. 27). For instance, the question 'How does the profitability of your [company/establishment] compare with that of the competitors in the market?' was rephrased as 'In 2018, did this establishment make a profit?'.

Questionnaire development and content

The ECS 2019 questionnaire was developed using the ECS 2013 questionnaire as a starting point. Experts and stakeholders, including representatives of the European Commission, governments, trade unions and employer federations, then provided input to refine the questionnaire and to adapt it to a push-to-web survey. Expert meetings were held in November 2016 and May 2017; a background paper on skills utilisation was prepared (Eurofound and Cedefop, 2020a); input from the OECD was collected to develop a module on skills; and experts provided item-by-item feedback.

The ECS 2019 questionnaire covers multiple topics. Figure 2 shows the conceptual framework which links the different topics. Two perspectives on the same topic are provided by interviewing a manager (MM) as well as (if present) an employee representative (ER).

Figure 2: The conceptual framework



Source: see Appendix 2

The MM questionnaire counts approximately 111 items covering 25 topics (Table 3). Out of 111 items, 28 items were also included in the ECS 2013, while four questions come from the OECD (three questions on skills mismatch and one on recruitment). Topics included in both waves are, amongst others, innovation, collective bargaining, teamwork, employee involvement and workplace involvement. The ECS 2019 also introduced new topics such as skills (mis)match, digitalisation and management strategies. The median survey time of the MM interview ranged from 22 minutes in the Netherlands to 30 minutes in Cyprus (Eurofound and Cedefop, 2020e, p.12).

Table 3: Topics in the MM questionnaire

	Topic	Number of questions	Questions from the ECS 2013
1	Characteristics of the establishment	3	3
2	Outsourcing and collaboration	5	2
3	Innovation	3	3
4	Staff numbers	1	-
5	Employee characteristics	3	2
6	Teamwork	3	3
7	ICT use	6	-
8	Hierarchy	5	1
9	Motivational levers	4	-
10	Job complexity and autonomy	3	-
11	Skills match	3	-
12	Skills development	3	-
13	Learning modes	3	-
14	Training and learning	8	-
15	Recruitment	7	1
16	Pay	4	-
17	Social dialogue	1	1
18	Collective bargaining coverage	6	6
19	Presence of a body for employee representation	4	3
20	Employee involvement	5	5
21	Employee influence	12	2
22	Workplace wellbeing	4	4
23	Workplace relations	2	-
24	Product market strategy	11	-
25	Respondent characteristics	2	2
	Total number of questions	111	38

Source: Own compilation based on Appendix 2

Because of the switch from telephone interviewing to online self-completion, it is difficult to compile credible time trends using the ECS 2013 and the ECS 2019 even when the same question is included in both waves (Eurofound and Cedefop, 2020a, p. 12). The reason is that one cannot tell whether

Disclaimer: This working paper has not been subject to the full Cedefop/Eurofound evaluation, editorial and publication process.

changes over time are due to a 'true' change or to the change in survey mode (also see section 3.2). Given that the consistency over time was in any case limited, EF/CF decided to re-design the questionnaire and re-translate all questions, even the questions that were already included in the ECS 2013. In this process care was taken to address issues with questions in the ECS 2013 (e.g. difficult concepts were clarified) and to make the questions suitable for an online survey.

Questions that were already included in the ECS 2013 are therefore often included in a different order in the ECS 2019 and are sometimes phrased differently. For example, the key question on establishment size in the ECS 2019 reads as: "***Approximately how many people work in this establishment³? Please include all people that are employed by the establishment, regardless of the type of contract and regardless of whether they are physically present or carry out their work outside of the premises. Each employee is counted as one person, regardless whether they are working full-time or part-time. Your best estimate is good enough.***" In the ECS 2013 this question reads as: "***Approximately how many employees work in this establishment? Please include all employees that are formally based in this establishment, regardless of whether they are physically present or carry out their work outside of the premises. Each employee is counted as one person, regardless whether they are working fulltime or part-time (= headcount). Your best estimate is good enough.***" The differences between the ECS 2013 and the ECS 2019 question are underlined.

While the two questions intend to measure the same concept, the ECS 2019 questionnaire refers to 'people' rather than 'employees' and clarifies that all people, regardless of contract type, should be included. Whether this slight difference in wording as well as the different order of the questions in the questionnaire, affects the responses and compromises the compilation of time trends is unclear. The survey literature has, however, shown that seemingly irrelevant changes in the question wording (de Bruin et al., 2012; Scherpenzeel & Saris, 1997) and question order (Van de Walle & Van Ryzin, 2011) can have substantial effects on the responses.

The ER questionnaire contains 101 items covering 14 topics (Table 4). Thirty-three items are related to employees' involvement in the decision-making process. The median survey time ranged from 16 minutes in Cyprus to 29 minutes in Malta (Eurofound and Cedefop, 2020e, p.28).

The ER and MM questionnaires contain many new questions in line with the conceptual framework (see Figure 2). An Excel file is available that gives the source of each question, the question's objective and rationale and defines complex (sub)concepts. For instance, the objective of the question '*What percentage of employees have the skills that are about right to do the job?*' is to 'assess skills match, under- and overskilling'. The rationale of this question is explained as 'the level of skills (mis)match is an important 'outcome' variable, which can be used both as an independent and a dependent variable'.

We believe that this approach could be strengthened by developing detailed research plans. Such research plans could explain how the questions will contribute to academic knowledge and/or inform policies (Groves et al., 2011; Rossi, Wright, & Anderson, 2013). Having such documentation publicly available, would also contribute to increasing the use of the data (see section 2.5). For instance, prior to the implementation of the survey, a research plan outlining which (policy relevant) research questions will be addressed with the new questions on skills (mis)match could have been developed. Shortly after the survey, this detailed research plan could then be implemented. We

³ Establishment is explicitly defined as: '*When the term 'establishment' is used in this survey, it refers to the local site at which you work.*'

therefore **recommend to document in future waves why (new) questions are included, how they will inform the debate and how they complement other sources and to make this documentation publicly available.** The European Social Survey, which already does this to some extent, can serve as inspiration.⁴

Table 4: Topics in the ER questionnaire

	Topic	Number of questions
1	Characteristics of ER and establishment	9
2	Resources	6
3	Communication	7
4	Employee support for ER	2
5	Information provision	10
6	Skill recognition and development	8
7	Digitalisation	2
8	Involvement in decision making	33
9	Involvement in determination of pay	5
10	Attitudes toward involvement	3
11	Relation with management	3
12	Work climate	3
13	Industrial action	9
14	Respondent information and follow up	1
	Total number of questions	101

Source: Own compilation based on Appendix 2

In future waves – and assuming that the survey mode does not change - **it is recommended to keep, to the extent possible, the wording and order of questions unaltered across waves so that the responses are comparable over time.** We support the suggestion of Eurofound (2017a, p. 85) **to identify a core set of questions that are included in each future wave.** This core set of questions can then be complemented with ad-hoc modules. Such an approach is standard practice in most other European surveys such as the EWCS and the ESS. It would increase the survey’s relevance as it would allow to study trends.

Defining concepts and the use of validated scales

The concepts highlighted in the tables above are captured by one or several questions. These questions need to be carefully worded so they are easily and correctly understood by the respondent. Significant efforts were invested in providing **clear definitions of key concepts** such as ‘establishment’, ‘team’ and ‘managers’.

The switch from a telephone survey to a web-based survey offers new options to present and clarify concepts and engage respondents (Tourangeau, Conrad, & Couper, 2013). One example is the use of

⁴www.europeansocialsurvey.org/methodology/ess_methodology/source_questionnaire/source_questionnaire_development.html.

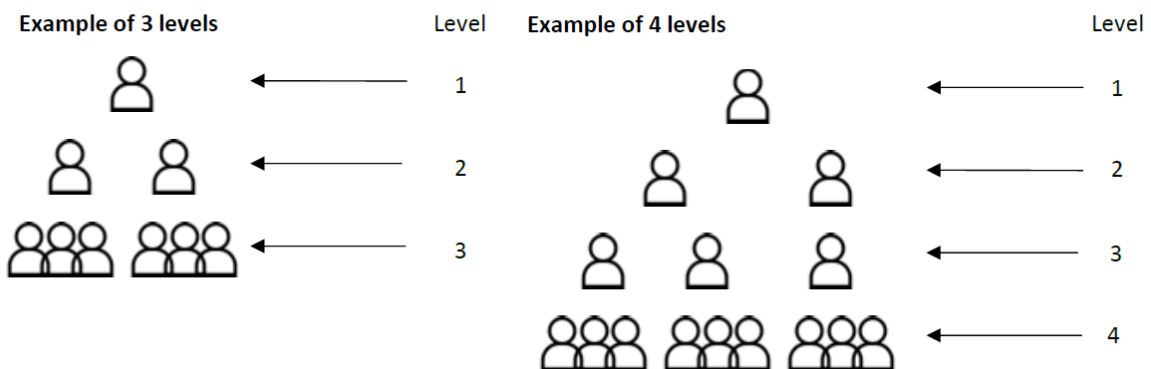
information buttons which provide definitions when respondents scroll over or click on a button. Such buttons were included in several questions to define, for example, the meaning of ‘meetings’.

Another interesting example of the new opportunities offered by web surveys is the use of pictures to clarify concepts (Couper, Tourangeau, & Kenyon, 2004). This option was used once to clarify the concept of ‘hierarchical’ levels in the MM questionnaire (Figure 3). Previous waves included similar questions. For instance, the ECS 2013 includes the question: “How many hierarchical levels do you have in this establishment, including the highest and the lowest level?”. Cognitive interviewing and the pilot test revealed that most respondents correctly interpreted the picture and provided correct responses (Eurofound and Cedefop, 2018a, d).

Figure 3: Example of including a picture to clarify a concept

25. [HIERA] Counting in the same way as is done in the examples below, how many hierarchical levels do you have in this establishment?

Indicate the number of hierarchical levels.



Source: Appendix 2

Survey research recommends using **validated scales** to measure complex concepts. Scales can be validated by testing whether they perform equally well in different contexts (e.g. in different regions, for different respondents ...). Once a scale is validated, responses can be compared across regions, across socio-economic groups and over time (Vander Elst, De Witte, & De Cuyper, 2014). The ECS measures complex concepts using scales, but it is not well documented whether these scales have been used before in cross-country research and have been validated. For instance, (autonomous) team work is captured by three questions in the ECS.⁵ These questions differ to a certain extent from the questions used in the EWCS to measure team work.⁶ **It would be useful to document whether this is a validated scale to measure team work or whether this scale still needs**

⁵ (1) "A team is a group of people working together with a shared responsibility for the execution of allocated tasks, within or across units of the establishment. What percentage of non-managerial employees at your establishment work in such a team?"; (2) "With regard to the non-managerial employees doing teamwork, do most of them work in a single team or do most of them work in more than one team at the same time?"; (3) "If you think about the tasks to be performed by the teams: Do the team members decide among themselves by whom the tasks are to be performed, or is there usually a superior distributing the tasks within the team?"

⁶ (1) "Do you work in a group or team that has common tasks and can plan its work?"; (2) "Is that always in the same team or group, or in several teams or groups?"; (3) For the team in which you work mostly, do the members decide by themselves...? [A. ... on the division of tasks; B. ... who will be head of the team; C. ... the timetable of the work]

to be validated, and whether other validated scales exist. EF/CF could consider providing funding to validate scales using the ECS data and consider compiling a publicly available dataset of validated scales. These scales could then be implemented in all EF/CF surveys (and, possibly, also in national surveys not funded by EF/CF) so that results from different surveys are comparable.

One new concept introduced in the ECS 2019 was **skills utilisation**. A background paper identified two conceptual approaches to measure skills utilisation (Eurofound, 2019). The first approach measures skills (mis)match. The second approach measures ‘enablers’, which captures workplace practices put in place to overcome skills mismatch. The ECS 2013 already included several questions related to the enablers approach (e.g. questions about training, motivation, autonomy). To measure skills mismatch, the ECS 2019 included the questions “*What percentage of employees have the skills that are about right to do the job?*”; “*What percentage of employees have a higher level of skills than is needed in their job?*”; “*What percentage of employees have a lower level of skills than is needed in their job?*”. As pointed out in the background paper, similar questions are already included in other surveys. Given that this type of questions is still relatively new, it would be useful to evaluate whether they are sufficiently informative (e.g. how do they relate to other ECS questions; how do they relate to indicators from other surveys measuring skills such as PIAAC) in order to decide whether these questions should be kept in future waves.

Questionnaire translation

Once the English source questionnaire was developed, the questionnaire was translated into the different target languages. This is a critical step in comparative cross-national research as all concepts should be interpreted in the same way across cultures and regions.

As the translation report explains (Eurofound and Cedefop, 2020c), a well-translated questionnaire meets three conditions: *semantic equivalence* across languages (the words and sentence structure in the translated text express the same meaning as the source language); *conceptual equivalence* across cultures (the concept being measured is the same across different cultures, although wording to describe it may be different); and *normative equivalence* to the source survey (the ability of the translated text to address social norms that may differ across cultures).

In our view, the translation process followed state-of-the-art procedures and met these three conditions. We first briefly describe the translation process (more details can be found in the cognitive interview report, the translation report and the technical and fieldwork report). Next, we provide a few examples that illustrate how the checks and balances ensured consistent translations and minimized translation errors.

As a first step of the translation process, EF/CF commissioned an advance translatability assessment of the draft source questionnaire in French, German and Polish (Eurofound and Cedefop, 2020a). A translatability assessment evaluates the extent to which concepts and measures can be meaningfully translated. It is considered a best practice as it helps to identify items that are difficult to translate so that these items can still be adapted in the source questionnaire (Conway, Acquadro, & Patrick, 2014).

In a second step, 9 cognitive interviews were conducted in the UK, Germany, France and Poland to test cognitive issues in different cultural and linguistic environments. These interviews focused on 19 questions in the MM questionnaire and 11 questions in the ER questionnaire and tested whether respondents correctly understood the questions and could easily provide correct answers. Based on

insights from the cognitive interviews as well as the translatability assessment, a few questions were revised. After the cognitive testing, a questionnaire finalisation meeting was held in March 2018 which resulted in the final English source questionnaire.

In a third step, the source questionnaire was translated to the target languages using the TRAPD approach (Translation, Review, Adjudication, Pretesting, and Documentation), which is considered the state-of-the-art approach to translate questionnaires for multi-country surveys (Harkness, 2007). For each target language, the English source questionnaire was first translated to the target language by two translators working independently. The differences were then reviewed by the adjudicator (a linguist) and discussed with the two translators during a review meeting. Based on this discussion, a final version was compiled. For target languages used in several countries (e.g. French in Belgium, France and Luxembourg), the final version for each country was compared and, when necessary, harmonised while preserving the local style of the language. For languages spoken in multiple countries but without major differences in the dialect (for example Greek in Greece and Cyprus), an 'adaptation' approach was followed, whereby the questionnaire was first translated for the country with the greatest number of speakers and then adapted for use in other countries. The full questionnaires in all languages have been made available on Eurofound's website. The questionnaire was then tested in a pilot study and a few questions were slightly revised (e.g. including an introduction to a question, revising scales) (Eurofound and Cedefop, 2018a, p. 119-121).

This translation processes ensured the highest quality standards. We give a few examples of (translation) issues that were detected along the way and corrected before mainstage fieldwork:

- The translatability assessment and the cognitive interviews revealed the need to use the term 'employee' consistently throughout the questionnaire, rather than referring to 'staff' or 'workers' (see IPSOS (2018, Table 1.5 & 1.6) for an overview of recommended changes by question).
- Based on feedback from translators and a check of EF/CF of the initial adjudicated version of the translation, a list of cross-country issues was compiled. Translators were then asked to check whether all issues mentioned on the list were correctly translated. For instance, translators verified whether 'skills' was correctly translated, and was not wrongly translated as 'qualifications' (see Annex 7 of the translation report for the full list of additional checks).
- The cross-nation review meeting to harmonise the questionnaires from Croatia, Montenegro and Serbia substantially improved the Croatian translation. For instance, the translation of 'establishment' was changed in the Croatian questionnaire.
- Eight generic ER types were identified across the EU28 Member States. The ER types are country specific and each country uses country-specific terms. For instance, the translation of the ER types for the Netherlands differ from those for Flanders, the Dutch-speaking region in Belgium. These terms were carefully translated so that respondents identified the correct ER type (see Annex I in the technical and fieldwork report for an overview of ER types by country).

During mainstage fieldwork one translation issue was detected in Sweden. Due to a translation mistake in the local language script, the NACE sectors were not correctly sorted in the MM questionnaire. As a result, respondents selected the 'no answer' or 'don't know response' and could

no longer answer the remaining questions. This substantially reduced the number of completed interviews in Sweden.

The translation report contains some minor practical recommendations so that the translation process will run even smoother in future waves. One suggestion we fully support is avoiding that the source questionnaire is amended after the translation starts as this creates additional work and may lead to different versions of the questionnaire in different countries. Ideally, amendments are to be avoided, but, if strictly necessary, then it is preferable to implement them all at once.

Revisions to the questionnaire after the pilot

One objective of the pilot was to test and validate the questionnaire. This allowed testing whether the questionnaire was not too complex, whether the respondents correctly understood the questions and whether some questions suffered from item non-response or straight lining. Importantly, the pilot included debrief questions so that respondents could indicate whether they experienced problems with certain questions. We give a few examples of insights gained from the pilot:

- The hierarchy question (see Figure 3) was relatively well understood by the respondents and was not modified.
- For questions asking to report the number of employees (e.g. the number of managers, the number of employees with open-ended contracts), the pilot showed that providing respondents with the option to report by either reporting a number or selecting a percentage category from a scale improved data quality.
- The questions on skill (mis)match were not sufficiently clear. To remedy this issue, an introduction was added, respondents were given the option to report a number or to select a percentage from a scale, and the introduction to the questions explicitly stated that the responses to the three items should add up to 100%.
- Due to a high level of item non-response to the question asking the ER to report the contractual working time in hours per week, a scale was added to this question.
- To improve the flow of the screener interview in countries with a company-level frame, interviewers first collected contact details of the MM and, if present the ER, before sampling establishments.

Qualitative evidence from mainstage fieldwork confirmed that the questionnaire was not too complex and well understood by the respondents. First, in all countries, the median survey length was at most 30 minutes. Second, straight lining and warning messages (e.g. when respondents entered implausible values) were rather exceptional. Third, respondents could contact Ipsos by email if anything was unclear, but Ipsos received few requests for help.

2.2. Sampling strategy, sampling frames and reference statistics

The ECS aims to be representative of the population of establishments employing at least 10 people in each of the countries covered, including the NACE sectors B to N, R and S and excluding the sectors A, O, P, Q, T and U. Importantly, and in contrast to the ECS 2013, the public sector was excluded from the universe.

Even more so than in household surveys, designing an adequate sampling strategy in company surveys is a challenging endeavour. In many respects, the ECS 2019 followed a sophisticated sampling strategy that managed to overcome many of the challenges inherent to company surveys and addressed several of the shortcomings identified in the ECS 2013 (Eurofound, 2017b). The sampling strategy is well documented in the sampling report, the weighting report and the technical and fieldwork report. We do not repeat this information here and only re-iterate key elements of the sampling strategy:

- Like the ECS 2013, the ECS 2019 used a random probability sample, explicitly stratified by size and sector (nine cells) and implicitly by top level NACE code and region.
- If possible, establishment-level register data were used.
- In order to maximise coverage, establishments without size information were included in the sample in five countries (the 'no size stratum'). As such, establishments that were by design excluded in previous waves now have a positive probability of being included in the sample. In addition, register entries without phone number were also included in the sample using the Dun & Bradstreet register to look up phone numbers. For these reasons, coverage in the ECS 2019 is higher than in the ECS 2013.
- To limit the proportion of remaining sample in cells for which the targets have been reached, fieldwork was closely monitored, and fully stratified batches were incrementally released when needed. Ipsos aimed to limit the proportion of live sample at the end of the fieldwork to less than 10% (Eurofound and Cedefop, 2020a, p.131).
- A new strategy was used to select establishments in multi-establishment sites for those countries where a company-level frame was used. This approach was designed to avoid under-representing multi-establishment sites (as in ECS 2013) and to avoid over-representing headquarters of multi-establishment sites (Eurofound and Cedefop, 2020a, p.36).

Table 5 summarises key elements of the sampling frames. The sampling frames are listed in Table 19 in Appendix 1.

Table 5: Sampling frame characteristics

Sampling frame characteristics	Number of countries (out of 28)
Establishment-level sampling frames	11
Coverage (excluding 'no size stratum')	
>90%	17
80% - 90%	7
<80%	4
Use of no size stratum	5 (Austria, Cyprus ⁷ , Greece, Luxembourg and Malta)
Phone numbers >90%	24

Source: Table 5 (Eurofound and Cedefop, 2020d)

⁷ In Cyprus, a no size stratum was used because the national statistical authority was not permitted to share the size of the companies.

As in previous ECS waves, it proved difficult to obtain complete and up-to-date establishment-level sampling frames. In some countries, establishment-level frames do not exist; in others the frame's quality is not sufficiently high. In 17 out of 28 countries establishment-level registers were not available, and company-level registers were used instead. This is an improvement compared to the ECS 2013, which used establishment-level sampling frames in 14 out of the 28 EU Member States (Vila & Carausu, 2013, Table 3). In 17 countries, the sampling frame covered at least 90% of the population. With some exceptions, all sampling frames were updated within a year preceding fieldwork. Coverage was lower than 80% in four countries (Austria, Greece, Luxembourg and Malta). In order to increase coverage in these four countries, the 'no size stratum' was included in the sampling frame. This raised coverage to over 90% in Austria and Luxembourg, to 77% in Malta and to 53% in Greece (Eurofound and Cedefop, 2020d, Table 7). This substantial increase in coverage proves the added value of including a 'no size stratum' in the sampling frame. In general, most sampling frames contained phone numbers of more than 90% of the companies/establishments. Ipsos matched the sampling frames with phone numbers obtained from Dun & Bradstreet which further increased phone number coverage.

While the sampling strategy was solid, pilot and mainstage fieldwork did encounter challenges due to the lower than expected yield rates. In other words, more establishments had to be contacted to achieve the targets than initially expected. The yield rates will be discussed in depth in the next section.

The challenges inherent to company surveys are reflected in the QA indicators (Table 6). Only five out of the eleven targets related to the accuracy of the sampling plan, frame development and reference statistics were achieved for the ECS 2019. In our view, the impact of missing the QA indicators 1.20, 1.3, 1.4, 1.12, 1.13, 1.14 and 1.27 is limited because these targets were almost reached (see the quality control report for details). For instance, the sampling frame covered at least 95% of the population in only half of the countries, but – as discussed above - the coverage exceeded 90% in 26 out of the 28 countries. However, the fact that the number of completed interviews is lower than initially planned (QA indicator 1.29) does have an impact on the data quality, particularly because the targets were already reduced in five countries after the pilot. A related quality issue, which is not immediately apparent from the QAP, is the disappointingly low number of ER interviews. Both issues will be discussed at length in the next sections.

During the weighting process, an error was detected in the Slovenian sample. The sample provider erroneously omitted companies in the service sector, implying that 19% of the intended target population in Slovenia was not included in the sample. As a consequence, the Slovenian sample is biased. Fortunately, this bias has only a limited impact on the EU estimates as Slovenia is one of the smaller EU countries (Eurofound and Cedefop, 2020a, p.34). While this is indeed true at the European level, researchers and policymakers that are primarily interested in Slovenia cannot rely on the ECS to obtain nationally representative statistics.

Reference statistics describe the population of establishments by sector and size. They are critical to assess coverage, to design the sampling and to calculate the sampling weights. In 17 out of the 28 countries, the reference statistics were available at the company level only. In those cases, reference statistics at the establishment level were estimated using the interview screener data.

Table 6: Accuracy indicators related to the sampling strategy and the sampling frame development

No.	Indicator	Target	Result	Assessment
Sampling strategy/plan				
1.1 (RQ)	A common sampling strategy is agreed that reflects current academic standards, as illustrated by academic references included in the sampling strategy document	Yes	Yes	✓
1.20 (RW)	Percentage of countries where the agreed variables are used for stratification	100%	88%	✗
Sampling frames				
1.3 (RW)	Percentage of countries where a sampling frame on the establishment level is used	43%	34%	✗
1.4 (RW)	Percentage of countries where the sampling frame covers at least 95% of the population	100%	50%	✗
1.5 (RW)	Percentage of sampling frame units that refer to non-existent or non-eligible establishments and companies	10%	6%	✓
1.6 (RW)	Percentage of countries where the sampling frame was updated within a year preceding fieldwork	100%	97%	✓/✗
1.7 (RQ)	Percentage of countries where specified information on stratification variables (sector and size class) is included in the sampling frame	100%	100%	✓
Reference statistics				
1.12 (RW)	Percentage of countries where reference statistics on the establishment level are used	43%	36%	✗
1.13 (RW)	Percentage of countries where the reference statistics fully cover the population	100%	87%	✗
1.14 (RW)	Percentage of countries where the reference statistics used for stratification were updated within a year preceding fieldwork	100%	93%	✗
1.15 (RQ)	Percentage of countries where specified information on stratification variables (sector and size class) is included in the reference statistics	100%	100%	✓
Net sample				
1.27 (RW)	Percentage of countries where the distributions across stratification categories of the net sample closely approximates the distributions of the universe (sampling plan) (deviations in the proportional size of each of the strata between the two should not exceed 5 percentage point)	100%	54%	✗
1.29 (RQ)	Percentage of countries where the net sample size \geq planned sample size	100%	57%	✗

Source: (Eurofound and Cedefop, 2020b)

Note: RQ: required indicator; RW: real world indicator

Other issues included lacking population statistics for particular sectors. In Croatia, for instance, population estimates for NACE K, R and S were not available (Eurofound and Cedefop, 2020d, p.6). The reference statistics were obtained from official sources such as the national statistical authority. It is, however, difficult to assess how the reference statistics are compiled and whether they are accurate.

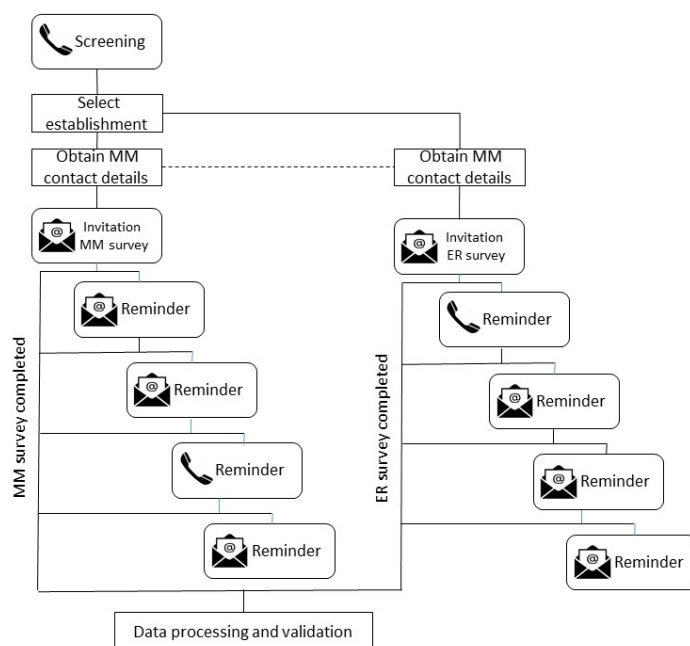
2.3. Fieldwork

To understand the quality of the survey processes as well as the data quality, the fieldwork organisation is first briefly discussed. The survey was carried out by Ipsos' country offices and coordinated by a Central Coordination Team (Project management lead, Femke De Keulenaer). Several strategies were put in place to assure and monitor data quality. These strategies are discussed at the end of this section.

A pilot study was set up in the 28 EU Member States and four candidate countries (North Macedonia, Montenegro, Serbia and Turkey). The pilot study aimed at testing the different aspects of the push-to-web methodology so that lessons learnt could inform mainstage fieldwork. It served at the same time as a full-dress rehearsal. Pilot fieldwork was conducted from the end of September until early December 2018 (see Figure 1 for a timeline). Pilot fieldwork for the screener interviews lasted between three and nine weeks. Experiences from the pilot informed mainstage fieldwork and led to the following changes in mainstage fieldwork (Eurofound and Cedefop, 2018a): the four EU candidate countries were dropped from the survey; the target sample size was revised downwards in six countries because yields were lower than expected; the best performing contact strategy for the screener interview (screener interview with senior manager, rather than, for instance, the receptionist) was selected; the timing and order of the reminders was optimised; the sampling strategy used to select an establishment of a multi-site company in countries where a company-level sample frame was used was revised; the relevance of inclusion of a 'no size stratum' to increase coverage was validated; and a few questions in the ER and MM questionnaire were revised. The pilot was an essential phase in the survey as it helped to test and refine the push-to-web methodology. The tweaks to the survey design and processes based on experiences from the pilot substantially improved mainstage fieldwork.

Mainstage fieldwork was carried out from January until July 2019. Figure 4 summarises the workflow. The first step - referred to as 'recruitment', the CATI (Computer Assisted Telephone Interview) stage or screening stage - is a telephone screener interview. This is a critical step in the survey as the interviewer (1) selects the establishment (only if a company-level register is used); (2) determines the eligibility of the establishment; (3) attempts to convince the gatekeeper and/or management to take part in the survey; and (4) obtains correct contact details (email address) of the manager and the ER (if present). In countries where a company-level register was used, establishments of multi-site companies had to be selected during the screener interview. This is a difficult step for interviewers as well as respondents.

Figure 4: Fieldwork



Source: Appendix 2

Based on experiences from the ECS 2013 and insights from the pilot, an elaborate and effective procedure was developed which (1) always selected the contacted establishment, (2) selected all eligible establishments in companies with at most three establishments and (3) randomly selected three eligible establishments in companies with more than three establishments (see EF/CF (2019f, p. 2 for details).⁸ To maximise response rates at the screener stage, six attempts were made to reach an establishment at different times and days of the week over a period of at least two weeks before discarding the establishment as a non-contact. The CATI interviews started on 23 January 2019 and the last screener interview was held on 20 June 2019 in Greece (see EF/CF (2019f, Section 10) for the exact dates of fieldwork in each country).

Once the MM and/or ER contact details are obtained, online questionnaires are sent (web survey). This is the CAWI (Computer Assisted Web Interview) stage of the survey. To increase response rates, the invitations were followed-up by three emails and one (or, in a limited number of cases two) phone reminders. In case of the MM interviews, a phone reminder was scheduled shortly after the second email reminder. The phone reminder for the ER interviews was scheduled shortly after the survey invitation was sent. This phone reminder was also used to collect the ER email address in the case that the MM had only shared the ER's phone number. The last MM and ER online surveys were completed on 25 July 2019.

The pilot played an important role to test different fieldwork strategies and, more specifically, to test different contact strategies at the screening stage. Four different contact strategies were tested (Eurofound and Cedefop, 2018a, p. 55 – 64). The key difference between these contact strategies

⁸ In companies with at most eight establishments, all establishments were listed and three establishments were randomly selected; in companies with more than eight establishments, the program randomly drew a letter and respondents were asked to select the establishment at a location starting with that letter (or the next nearest letter).

was whether the interviewer conducted the screener interview with the first contact person reached or attempted to recruit a (senior) MM who was then also asked to complete the online survey. This experiment revealed that it required more effort to recruit MM as respondents to the screener interview, but that the CAWI conversion rates were higher when a MM was interviewed. Overall, it turned out that recruiting a senior MM (rather than interviewing the first contact person) resulted in more completed MM questionnaires. Moreover, the data quality tended to be higher when the senior MM completed the questionnaire. Furthermore, management respondents were more likely to indicate the presence of an ER in the establishment (Eurofound and Cedefop, 2018a, p. 117). Mainstage fieldwork adopted the optimal contact strategy, and interviewers always attempted to recruit the senior manager at the screening stage. The pilot also helped to improve the strategy to obtain ER contact details and helped to fine-tune the timing of the email reminders.

Table 7 summarises the QA indicators related to the contact strategy, respondent follow-up and interview targets. Five out of seven targets were achieved. The missed targets are again related to the lower than expected yield rates, which explains why indicators 9.1 (% of countries where the used gross sample is smaller or equal to the planned gross sample) and 9.2 (% of countries where at least 90% of target was achieved) were missed. The following sections will consecutively discuss the achieved sample size and fieldwork progress, the MM and ER yield rates and data quality monitoring.

Table 7: Accuracy indicators related to the contact strategy, respondent follow-up and interview targets

No.	Indicator	Target	Result	Assessment
Contact strategy				
9.1 (RW)	Percentage of countries where the used gross sample is smaller or equal to the planned gross sample	100%	56%	✘
9.2 (RQ)	Percentage of gross sample entries that are discarded before the net sample is realised, for which a final outcome has not been realised (i.e. no cases to be lost)	100%	0%	✓
9.3 (RW)	Percentage of gross sample entries that are contacted at least once, but are discarded upon realising the net sample (i.e. cases that were opened but not finalised)	<10%	5%	✓
9.4(RQ)	Percentage of sample entries to which a final status of 'non-contact' was assigned that were not called at least six times	0%	0%	✓
Respondent follow-up (Lot 1)				
9.5 (RQ)	Percentage of respondents that agreed to participate but for which final non-response is accepted that were not reminded with at least three emails and at least one phone call	0%	0%	✓
Interview targets				
10.15 (RQ)	% of countries where at least 90% of target was achieved	100%	82%	✘
10.16 (RQ)	At least 95% of target across countries reached	100%	100%	✓

Source: (Eurofound and Cedefop, 2020b)

Note: RQ: required indicator; RW: real world indicator

Achieved sample size and fieldwork progress

In all countries, fieldwork started at the end of January 2019 and lasted between 14 and 21 weeks (Eurofound and Cedefop, 2020a, Table 10). The sample size targets of MM interviews differ by country, they range from 250 in small countries to 1,500 in larger countries (Table 8).

Table 8: Target sample size (mainstage fieldwork)

Number of countries	Countries	Target sample size (MM interviews)
4	Cyprus, Ireland*, Luxembourg, Malta	250
1	Slovakia*	350
6	Croatia, Estonia, Greece, Latvia, Lithuania, Slovenia	500
1	UK*	700
12	Austria, Belgium, Bulgaria, Czechia, Denmark, Finland*, Germany*, Hungary, Netherlands, Portugal, Romania, Sweden	1,000
1	Poland*	1,150
3	France, Italy, Spain	1,500

Source: Appendix 2

Note: * indicate countries where targets were reduced after the pilot

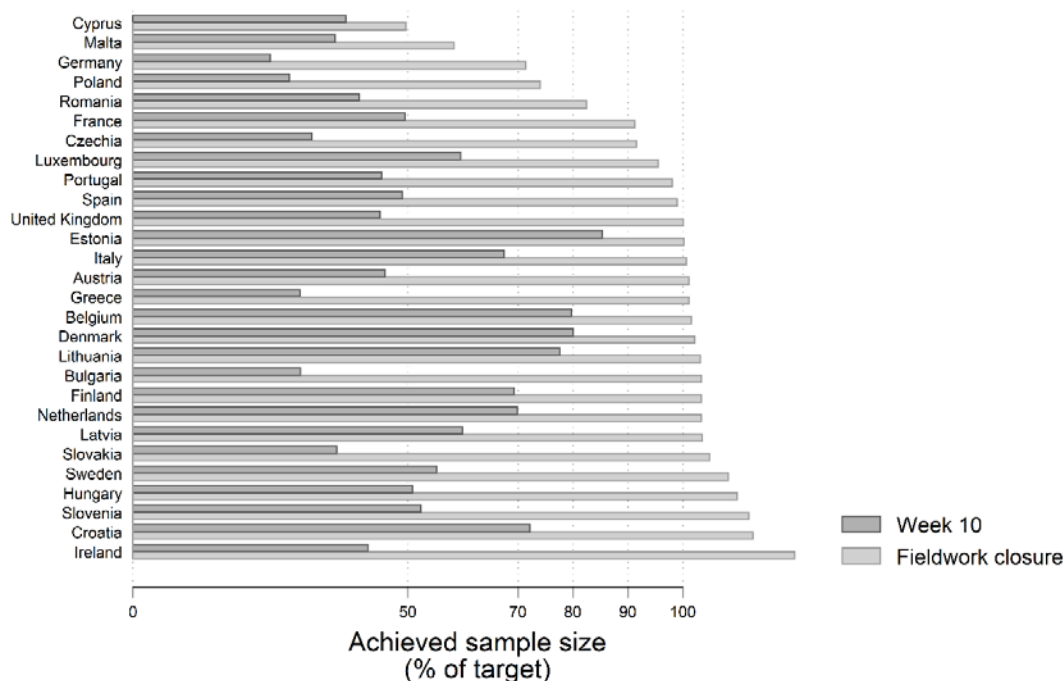
No targets were set for the ER interviews. Note that the sample size targets for mainstage fieldwork were reduced in Finland, Germany, Ireland, Poland, Slovakia and the United Kingdom after the pilot (Table 20 in Appendix 1). The pilot revealed that yield rates in those countries were too low to achieve the initially proposed targets. In Finland and the UK, the targets were increased again during mainstage fieldwork as the yields during mainstage fieldwork were better than expected.

Figure 5 shows the share of completed MM interviews relative to the target sample size by country. The target was achieved in 18 out of 28 countries by the end of the fieldwork. In five countries (France, Czechia, Luxembourg, Portugal and Spain) the sample size was within acceptable levels (>90% of target). In three countries (Germany, Poland and Romania) the achieved sample size was between 70% and 90% of the target, whereas in two countries (Cyprus and Malta) the achieved sample size was lower than 70% of the target.

Importantly, we cannot conclude from these figures that the local country offices that did not achieve the targets underperformed. In general, yield rates (the share of completed MM questionnaires to sample dialled, discussed in more detail in the next section) are lower in countries where the targets were missed. In other words, country offices that missed the targets had to reach out to substantially more establishments to achieve the same number of completed interviews than country offices that achieved the targets. Moreover, field visits by Cedefop and Eurofound in Germany, Greece, Ireland⁹, Italy and the UK during fieldwork did not reveal major problems that can explain the low achieved sample size.

⁹ The field visit in Ireland revealed some issues with, among others, the training of the interviewers. These issues were quickly addressed, and the targets were reached in Ireland.

Figure 5: Achieved sample size (MM) by week 10 and by closure of fieldwork



Source: Own compilation based on Table 18 (Eurofound and Cedefop, 2020a)

Note: The final sample size is slightly lower than the sample size at fieldwork closure because 161 MM interviews were of low-quality and removed from the dataset. Table 20 in Appendix 1 gives the target and final sample size by country

The achieved number of MM interviews is particularly low in Malta and Cyprus, where only 124 and 146 questionnaires were completed respectively. This is due to the low number of establishments in the sample frame. As a result, all establishments in the sample frame were contacted (i.e. a census was conducted) and Ipsos ran out of sample before the targets were achieved. Hence, in order to achieve the targets in Malta and Cyprus in future waves, higher yield rates need to be achieved in Malta and Cyprus than in other countries as it is not possible to release more sample in these two countries.

Figure 5 also gives an indication of fieldwork progress by country. We do not observe clear patterns. Some countries that were lagging behind after ten weeks of fieldwork, such as Bulgaria, Hungary and Ireland, managed to accelerate the process and still (nearly) achieved the targets by the end of the fieldwork. Other countries, such as Germany and Poland, were lagging behind after ten weeks of fieldwork and never caught up.

As shown in Figure 4, the MM survey invitation was followed-up by four email reminders and one phone reminder. **The phone reminder appears critical to improve the response rate:** 19% of the MM who were reached during the CATI reminder eventually completed the survey (Eurofound and Cedefop, 2020a, Table 52).

Respondents were also offered a customised report which benchmarked the respondents' workplace against similar workplaces on key indicators such as business performance and work climate. These customised reports are prepared by Eurofound and then shared by Ipsos (who adds the establishment's name) with the respondent. Interviewers positively evaluated offering such a customised report to the respondents. The share of respondents who asked to receive a customised report ranges from 37% in Poland to 80% in Finland. At the time of writing (March 2020), the

respondents had not yet received the customised reports. If technically feasible, it would be ideal if respondents received the reports a few weeks after the end of fieldwork.

In three countries, additional incentives were offered to the respondents. In Czechia and Slovenia a €5 donation was given to a charity for each respondent, whereas respondents received a €5.4 shopping voucher in Finland.

MM yield rates

An important driver of the cost of a survey are the yield rates, which is defined in this report as the number of completed questionnaires relative to sample dialled (Eurofound and Cedefop, 2020a, p. 68).¹⁰ The management yield rate is determined by (1) the share of successful screener interviews, i.e. the share of MM that were sent an online questionnaire relative to the sample dialled, and (2) the conversion rate, i.e. the share of MM that completed the survey out of all MM that received the online questionnaire. The conversion rate only plays a role in push-to-web surveys such as the ECS 2019.

Table 9 shows the MM yield rate by country. Across all countries, the average yield rate is 5.2%, ranging from 1.5% in Poland to 15.9% in Lithuania. The low yield rate results from a combination of a low proportion of successful screener interviews and a low conversion rate. The share of successful screener interviews is on average 16% (ranging from 5% in Poland to 50% in Lithuania), while the average conversion rate is 34% (ranging from 17% in Germany to 50% in Czechia and France). Both the proportion of successful screener interviews and the conversion rates are lower than initially anticipated. Prior to the pilot survey, the share of successful screener interviews was expected to be 33.3% to 50%, whereas the expected conversion rate was 33.3%, implying an overall yield rate between 11.1% and 16.7% (Eurofound and Cedefop, 2020d, p.15). After the pilot, these expectations were revised downwards as the conversion rates were lower than expected, which also led to a revision of the sample size targets.

Importantly, we do not observe a correlation between the share of successful screeners and the conversion rate (Figure 6). In other words, countries with a low proportion of successful screeners can have high conversion rates and vice versa. For instance, the share of successful screeners in Greece is high (31%), while the conversion rate is low (18%). By contrast, the survey in Czechia is characterised by a low share of successful screener interviews (6%) and a high conversion rate (50%). In Slovenia and Lithuania – the two countries with the highest overall yield rate – the share of successful screener interviews was particularly high (about 50%), while the conversion rates were close to the EU average (34%).

¹⁰ This is not the same as the 'response rate'. Different definitions of the response rate, the refusal rate and the cooperation rate exist (see AAPOR guidelines, <https://www.aapor.org/Education-Resources/For-Researchers/Poll-Survey-FAQ/Response-Rates-An-Overview.aspx>). In general, the response rate is defined as the ratio of eligible respondents that cooperate in a survey to the total number of eligible respondents. The yield rate does not distinguish between eligible and ineligible respondents, but simply counts the number of completed questionnaires to the issued sample. Yields rates are thus by definition lower than the response rate.

Note that our definition of the yield rate differs from the definition commonly used by Ipsos and EF/CF who define the yield rate as the number of questionnaires sent (rather than the number of questionnaires completed) to gross issued sample and/or sample dialled.

Table 9: The MM yield rate

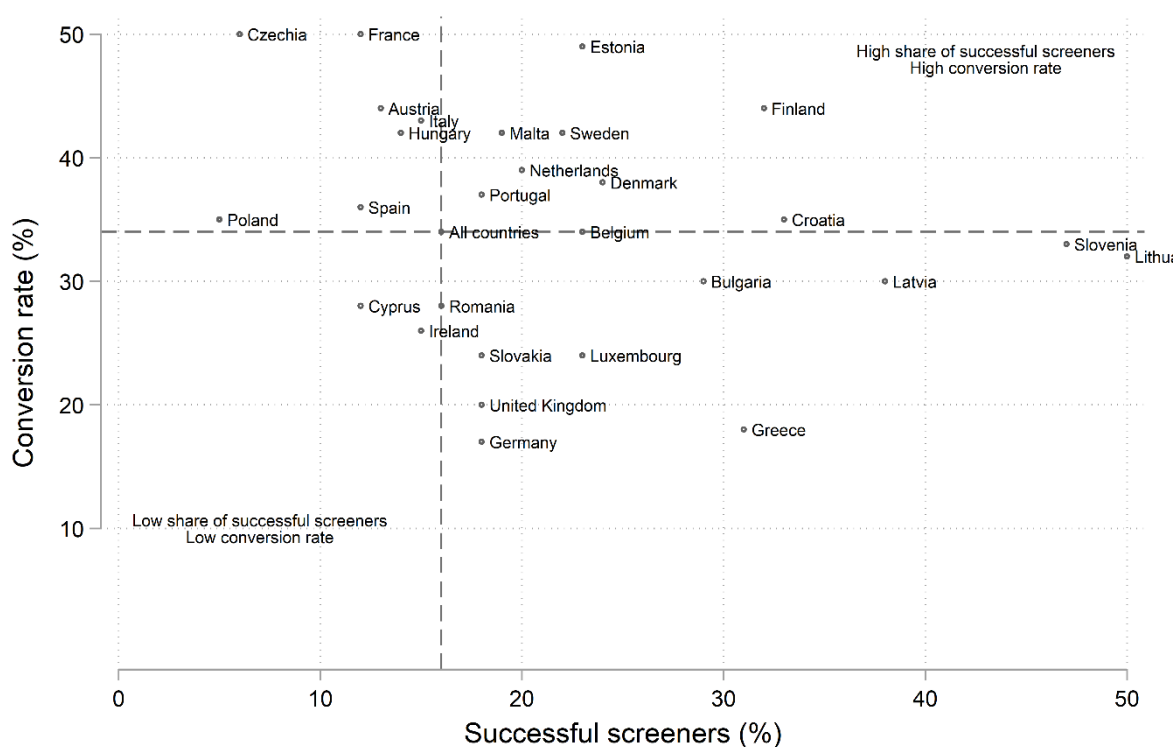
	Share of successful screeners to sample dialled	Conversion rate (=completed questionnaire/invitations sent)	The MM yield rate (=share successful interviews * conversion rate)
All countries	16%	34%	5.2%
Austria	13%	44%	5.6%
Belgium	23%	34%	7.5%
Bulgaria	29%	30%	7.7%
Croatia	33%	35%	11.1%
Cyprus	12%	28%	3.3%
Czechia	6%	50%	2.7%
Denmark	24%	38%	9.1%
Estonia	23%	49%	11.0%
Finland	32%	44%	14.2%
France	12%	50%	6.1%
Germany	18%	17%	2.9%
Greece	31%	18%	5.6%
Hungary	14%	42%	5.2%
Ireland	15%	26%	3.7%
Italy	15%	43%	6.2%
Latvia	38%	30%	10.7%
Lithuania	50%	32%	15.9%
Luxembourg	23%	24%	5.1%
Malta	19%	42%	8.1%
Netherlands	20%	39%	7.6%
Poland	5%	35%	1.5%
Portugal	18%	37%	6.4%
Romania	16%	28%	4.2%
Slovakia	18%	24%	4.3%
Slovenia	47%	33%	15.0%
Spain	12%	36%	4.2%
Sweden	22%	42%	8.4%
United Kingdom	18%	20%	3.5%

Source: Own compilation based on Table 21 and Table 22 (Eurofound and Cedefop, 2020a)

Figure 6 also reveals that the low overall yield rate in Poland (1.5%) is primarily caused by a low share of successful screeners (5%), while the conversion rate in Poland (35%) does not differ substantially from the EU average. **In order to improve overall yield rates in future push-to-web surveys and, hence, reduce costs, some countries will have to focus on improving the rate of successful screeners, while other countries should prioritise improving conversion rates.**

Discussions with the fieldwork contractor and the involved country offices as well as the social partners may help to devise country-specific strategies to improve the yield rate. In addition, it could be considered to offer incentives to respondents, as was done in Czechia, Slovenia and Finland. As discussed in section 4, in countries with particularly low conversion rates (e.g. Germany, Greece, and the UK) it might be considered to conduct a phone survey (as in previous waves) rather than a push-to-web survey.

Figure 6: Correlation between conversion rates and successful screeners: countries with low conversion rates do not necessarily have low levels of successful screeners and vice versa



Source: Own compilation based, see Appendix 2

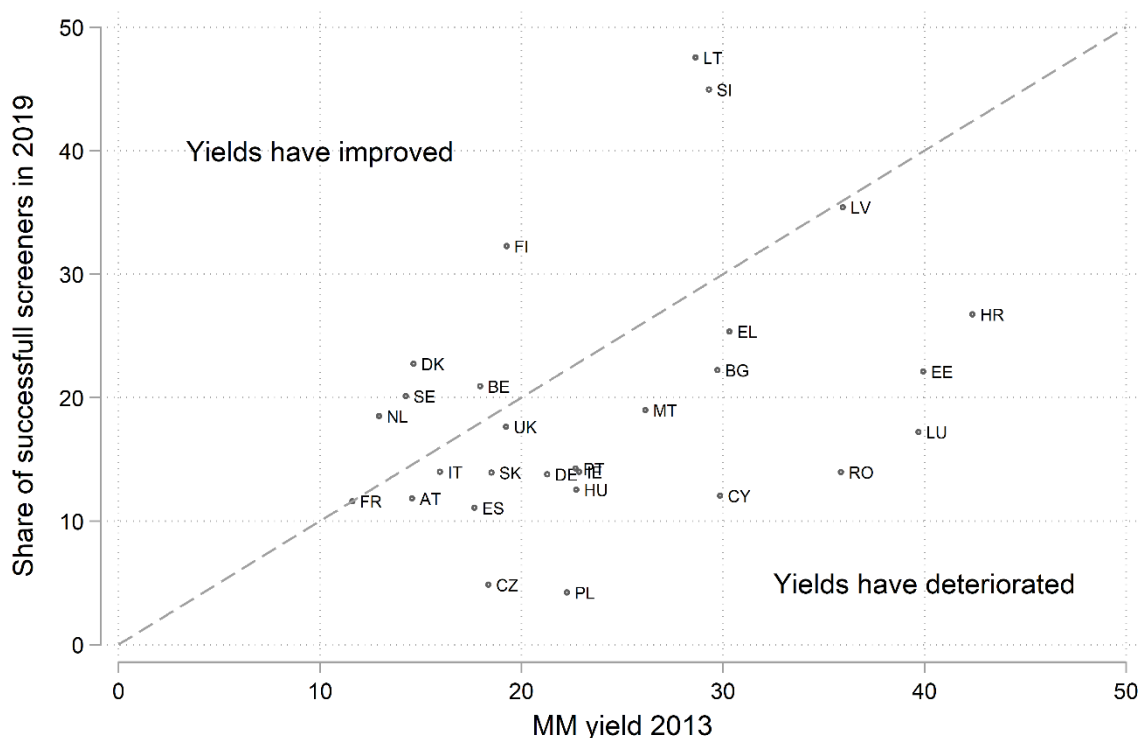
Note: The dashed lines indicate the average share of successful screeners and the average conversion rate across the EU28

It is interesting to compare the share of successful screener interviews in the ECS 2019 to the share of successful interviews in the ECS 2013 (Figure 7). As the ECS 2013 was a telephone survey, conversion rates do not play a role. The average yield rate across the EU28 was 19% in 2013, which is similar to the proportion of successful screener interviews (16%) in the ECS 2019. One explanation for these similar rates is that the screener interview is quite complex, particularly in multi-establishments companies. Another explanation is that for respondents who are willing to participate in a survey, it does not really matter much whether the survey takes 5 minutes (e.g. the 2019 screener) or 25 minutes (e.g. the 2013 phone survey). We do, however, observe quite some variation between countries. The correlation between yields in 2013 and the proportion of successful screener interview in 2019 is moderate (correlation: 0.43). **The findings suggest that yield**

rates obtained in previous waves do not always accurately predict yield rates in future waves, complicating the design of future waves.

Across the EU, the yield rate in 2013 (19%) and the proportion of successful screener interviews in 2019 (16%) have a same order of magnitude. This shows that the lower yield rate in the ECS 2019 is primarily driven by the implementation of a push-to-web survey. This requires managers - after they have been identified during the screener interview - to complete the online questionnaire. This additional step reduces yield rates as only about one in three managers completes the online survey. **Hence, the switch from a phone interview to online self-completion has contributed to the substantially lower yield rates in the ECS 2019 compared to the ECS 2013.**

Figure 7: Relation between proportion of successful screener interviews in 2019 and the MM yield rate in 2013



Source: Own compilation, see Appendix 2

A screener interview can be unsuccessful for several reasons (Eurofound and Cedefop, 2020a, Table 21). Across all countries, 42% of the establishments refused to collaborate (either because the respondent is not interested or because this is the company policy). This is similar as with the ECS 2013 when refusal-related issues accounted for 39.7% of the non-response in the management survey (Vila & Carausu, 2013, p. 39). About one in ten establishments could not be reached due to a 'wrong' phone number.

Another concern related to the quality of the screener interviews is the sample that remains 'open' or 'live' after fieldwork closure. Open sample consists of establishments that are part of the sample, but which have not been contacted (at least six attempts) before the end of the fieldwork. For instance, in many surveys the remaining establishments in a particular cell (e.g. establishments of the same size and sector) will no longer be contacted when the target within this cell is reached. Open sample can lead to bias as firms that have not been reached may differ from firms that have been reached. Ipsos proposed an innovative strategy to limit open sample (Eurofound and Cedefop,

Disclaimer: This working paper has not been subject to the full Cedefop/Eurofound evaluation, editorial and publication process.

2020a, p. 131). Sample was released in batches of establishments within the same cell when (weekly) projections showed that the targets would not be reached within this cell. This strategy avoided that additional sample was released in cells where the target could be reached without releasing additional sample and minimised open sample at fieldwork closure. The remaining open sample was 4.6% across all countries and only higher than 10% in Italy (Eurofound and Cedefop, 2020a, Table 57). In the ECS 2013 14% of the sample remained open, which shows that releasing fully stratified batches worked well to limit open sample. **Hence, this strategy should also be adopted in future waves.**

ER yield rates

Up to now the focus was on yield rates with respect to the management survey. A similar indicator can be derived for the ER interviews. During the screener interviews, the interviewer verified whether an employee representative is present and attempted to obtain an email address of the ER from the MM. ER contact details were also collected in the MM survey. The total number of completed ER interviews thus depends on (1) the share of successful screener interviews; (2) the presence of an ER in the establishment; (3) obtaining a valid email address in the screener interview or the MM interview; and (4) the conversion rate, that is the number of completed ER questionnaires to the total number of ER questionnaires sent. Like the MM yield rate, the ER yield rate is defined as the number of completed ER questionnaires to the number of establishments contacted (i.e. sample dialled).

The ER yield rate is, by definition, lower than the MM yield rate because many establishments are relatively small and may, therefore, not have an ER. In the ECS 2013, an ER was present in 32% of the establishments in the EU28, but with large differences between establishments in different countries, in different sectors and in establishments of different sizes (EF, 2015, Figure 44). In addition, even if an ER is present, the MM is not always willing to share the ER contact details, which further reduces the number of ER interviews that can be achieved.

Table 10 shows the different components of the ER yield rate. 3,095 ER interviews were completed, which is only a fraction of the 22,030 completed MM interviews. In only eleven countries (Austria, Belgium, Bulgaria, Denmark, Finland, France, Italy, Lithuania, Netherlands, Spain and Sweden) were more than 100 ER interviews completed. The number of completed ER interviews to the total number of sample dialled (the ER yield rate) is substantially lower than the MM yield rates. The average ER yield rate is 0.7%, ranging from 6.4% in Finland to less than 0.1% in several countries.

The low number of achieved ER interviews relative to the number of achieved MM interviews results from a combination of a low (reported) presence of employee representation structures within the establishments and a low share of valid email addresses obtained in establishments with an employee representation structure. According to the screener interviews, an ER is present in 35% of the establishments (column 4, Table 10). In 30% of those cases, valid ER contact information was obtained, and an ER questionnaire was effectively sent (column 5, Table 10). **During the screener interview, 68% of the managers refused to provide the contact details of the ER (if present) and just 24% provided the name, phone number and email address of the ER (Eurofound and Cedefop, 2020a, p. 85).** In addition, the majority of the MM who broke off the screener interview did so at the section about the presence and contact details of the ER. The refusal rate is substantially higher than in the ECS 2013.

Table 10: The ER yield rate

	Questionnaires completed	Share of successful screeners (out of sample dialled)	ER present (out of screeners completed)	ER invitations sent (out of screeners completed)	ER invitations sent/ER present	Conversion rate (=completed ER questionnaire/invitations sent)	ER Yield rate
All countries	3,095	16%	35%	11%	30%	44%	0.7%
Austria	166	13%	32%	14%	44%	52%	0.9%
Belgium	100	23%	45%	15%	32%	23%	0.8%
Bulgaria	101	29%	31%	7%	22%	41%	0.8%
Croatia	72	33%	30%	10%	33%	45%	1.5%
Cyprus	3	12%	14%	2%	14%	33%	0.1%
Czechia	46	6%	16%	5%	34%	43%	0.1%
Denmark	134	24%	55%	13%	25%	37%	1.2%
Estonia	17	23%	17%	4%	25%	38%	0.4%
Finland	470	32%	63%	37%	59%	53%	6.4%
France	429	12%	67%	29%	43%	54%	1.9%
Germany	67	18%	26%	5%	18%	31%	0.3%
Greece	9	31%	4%	1%	25%	32%	0.1%
Hungary	34	14%	14%	3%	20%	45%	0.2%
Ireland	6	15%	27%	5%	17%	11%	0.1%
Italy	188	15%	31%	10%	31%	54%	0.8%
Latvia	12	38%	5%	2%	40%	33%	0.3%
Lithuania	133	50%	52%	22%	42%	38%	4.1%
Luxembourg	39	23%	57%	11%	19%	33%	0.8%
Malta	5	19%	10%	4%	38%	38%	0.3%
Netherlands	341	20%	45%	26%	59%	48%	2.5%
Poland	48	5%	31%	5%	16%	37%	0.1%
Portugal	17	18%	8%	1%	16%	52%	0.1%
Romania	77	16%	63%	10%	16%	25%	0.4%
Slovakia	27	18%	24%	5%	22%	33%	0.3%
Slovenia	23	47%	12%	4%	30%	37%	0.6%
Spain	196	12%	51%	9%	18%	50%	0.6%
Sweden	309	22%	56%	20%	35%	56%	2.4%
United Kingdom	26	18%	31%	5%	15%	16%	0.1%

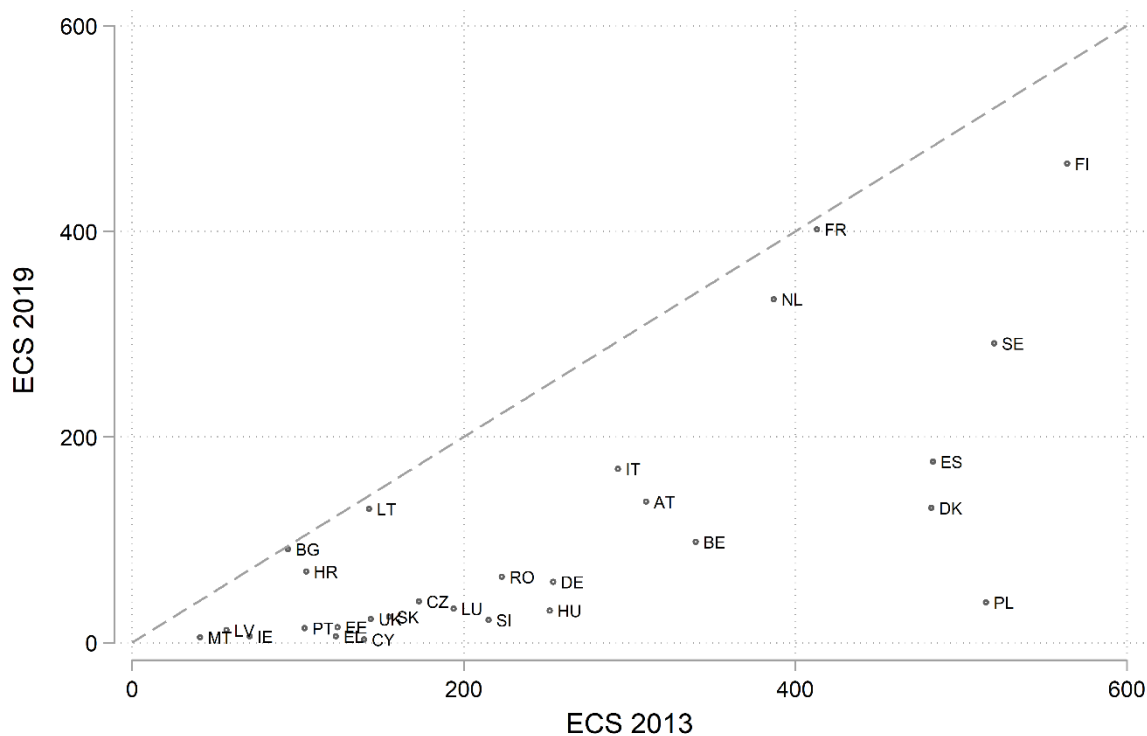
Source: Own compilation based on Table 21, Table 23 and Table 33 (Eurofound and Cedefop, 2020a)

Note: The final sample size is slightly lower than the number of completed ER questionnaires because 22 ER interviews were of low-quality and removed from the dataset

One potential explanation is that in 2013 MM were asked to provide ER contact details at the end of the phone interview. This interview may have established a sufficient level of trust between the MM and interviewer to convince the MM to share contact information. The introduction of GDPR at the time of mainstage fieldwork may also explain the MM's reluctance to share contact information such as email addresses.

The ECS 2013 interviewed 6,739 ER across the EU, excluding respondents from the public sector (Eurofound, 2015, Table 12, p. 146), compared to 3,095 ER in the ECS 2019. In each country, the number of completed ER interviews is lower in 2019 than 2013 and often by considerable margins (Figure 8). The most problematic case is Poland, where the ECS 2013 realised 515 ER interviews compared to only 48 interviews in the ECS 2019. **The very low number of completed ER interviews in the ECS 2019 implies that in most countries the sample size is too small for any meaningful analysis at the country-level, although aggregate EU level analyses might still be possible. This raises the questions whether (1) the ER yield rate can be increased in future waves to achieve similar sample sizes as in the ECS 2013 or (2) whether it should be considered to no longer conduct ER interviews as part of the ECS.**

Figure 8: The number of ER interviews in ECS 2019 and ECS 2013



Source: author calculations on ECS 2019 management dataset and Table 12 (Eurofound, 2015)

Note: To make the figures from ECS 2013 and ECS 2019 comparable, the ER interviews in the public sector in the ECS 2013 were not included. The figure gives the absolute number of ER interviews. Note that the target MM sample size differs by country and is not the same in both waves so that the number of ER interviews would in any case diverge between countries even if the proportion of successful ER to MM interviews were the same

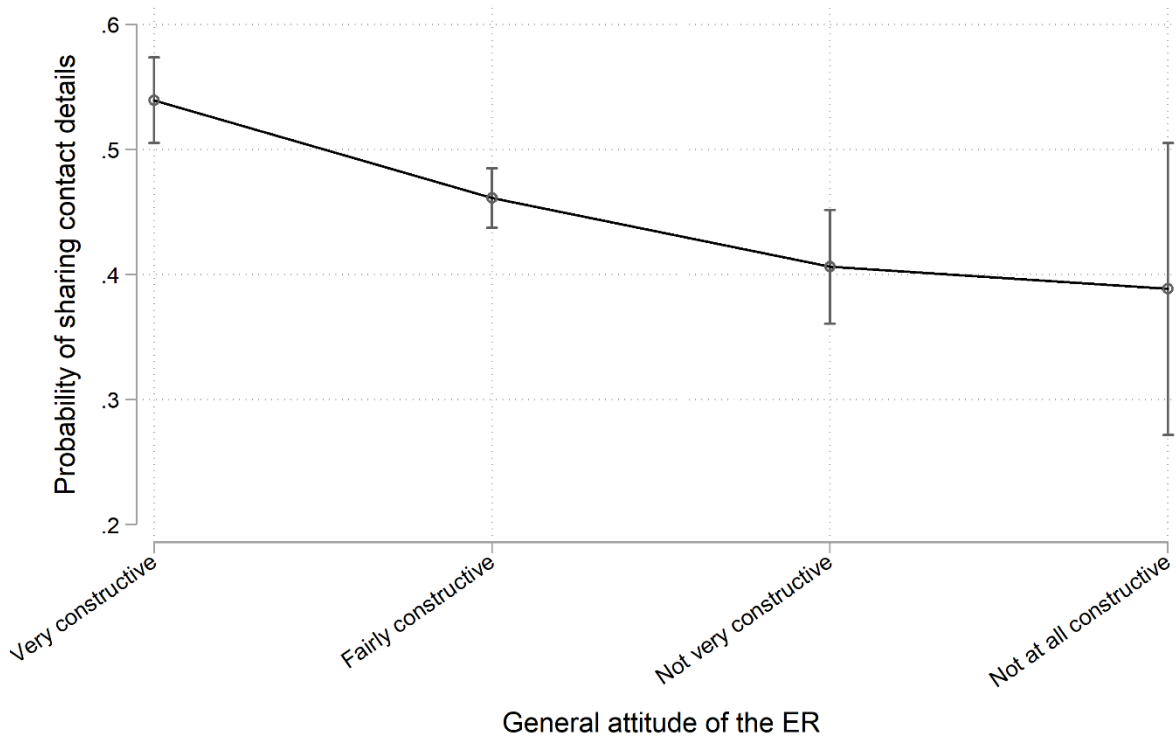
The low number of achieved ER interviews cannot be explained by insufficient efforts to reach ER. Throughout mainstage fieldwork, several additional actions were implemented to increase the number of achieved ER interviews. First, MM could provide ER contact details during the screener as well as the online interview. Second, if no email address was obtained during the screener interview nor during the MM online interview, an attempt was made to obtain the ER contact details by

Disclaimer: This working paper has not been subject to the full Cedefop/Eurofound evaluation, editorial and publication process.

phone. Third, a second phone reminder was conducted for ER respondents who were willing to complete the survey after the first phone reminder but had not yet done so. However, these additional actions only increased the number of completed ER questionnaires by 13% (from 2,730 to 3,095) (Eurofound and Cedefop, 2020a, Table 40). **It appears therefore unlikely that the number of ER interviews can substantially be increased in future waves by simply scheduling more (phone or email) reminders.**

Theoretically, in each establishment an interview with the MM and – if present – the ER could be conducted. In practice, the link between the ER and the MM is often lost as it happens frequently that only one (or neither) of them completes the online survey. The dataset contains only 1,835 establishments for which both the ER and the MM were interviewed (Eurofound and Cedefop, 2020a, p.133). Due to this low number of observations comparing responses of the ER and the MM within the same establishment, which we consider a unique selling point of the ECS, warrants care. The findings will not be representative for the EU28 because the number of observations is too small and there are no cases in a few countries.

Figure 9: Selection bias: MM with a good working relation with the ER are more likely to share the ER contact details



Source: author calculations on ECS 2019 management dataset and screener dataset

Note: Results from a logistic regression model that controls for firm size. Results are similar when the regression also controls for country. Full results are reported in Table 22 in the Appendix 1.

Finally, the approach used to obtain ER contact details may induce **selection bias**. The ER contact details are obtained through the MM. It is plausible that the MM with a good working relation with the ER are more likely to share contact details. This results in selection bias: interviewed ER differ systematically from those who were not interviewed. While it goes beyond the scope of this report to carefully assess selection bias and examine its impact on data quality, an explorative analysis indicates selection bias. Using logistic regressions, we examined whether managers who have a good

working relation are more likely to share the ER contact details. We find that 54% of the MM reporting that the ER have a very constructive attitude share the ER contact details (either an email address or a phone number), whereas only 41% of the MM reporting that the ER are not very constructive share the contact details (Figure 9, full results in Table 22 in the Appendix 1). These differences are statistically significant. It confirms that selection bias matters. As a result, the ER who are invited to participate in the survey will have a different profile than the ER who are not invited to participate, which leads to a biased ER sample. Selection bias is also a serious threat in previous waves, but may be more pronounced in this wave because 68% of the MM refused to provide ER contact details during the screener interview.

Monitoring data quality

Several strategies were put in place before, during and after the fieldwork to maximise data quality and limit non-sampling errors. The quality assurance plan contains thirteen indicators related to the accuracy of fieldwork training, scripting, fieldwork monitoring, data validation and data processing, of which eleven were met (Table 11). The first indicator that was missed concerns the number of (quality) issues that were detected during the weekly briefing and resolved. Out of the 33 issues detected, 29 were resolved.

The fact that several issues were detected and subsequently resolved indicates that excellent data monitoring processes were put in place. Issues with the sampling frame in Malta and Cyprus and two other issues in Poland and Germany could not be resolved. The second missed indicator concerns a scripting error in the pilot. This error did not affect mainstage fieldwork.

Avoiding non-sampling error starts with the design of the questionnaire and the interviewing process. Survey design and translation has already been discussed in Section 1. The interviewing process was also designed to reduce measurement error. One example is the strategy used to capture the sector of activity (Eurofound and Cedefop, 2020f). This is a key variable as it determines the eligibility of the establishment and the sampling weights. During the screener interview, the managers were asked to confirm the top-level (one digit) NACE code assigned to their establishment in the sample frame. If the respondent rejected the top-level NACE code, the interviewer asked for the correct sector. If an ineligible sector was selected, a warning message was shown to the interviewer to avoid that establishments were accidentally screened out. During the online MM interview, managers had to confirm the top-level NACE code and were then asked to select the next level (two digit) NACE code. This elaborate procedure differs from the ECS 2013 where an open-ended question was used to capture the sector. The ECS 2019 procedure seems better suited to avoid classification errors.

In order to harmonise fieldwork across countries, nearly all local project/country coordinators participated in a face-to-face fieldwork manager instruction meeting (QA indicator 7.1). Prior to the fieldwork, all interviewers participated in training and received comprehensive training materials and briefing notes (QA indicator 7.6 and 7.7). This training often included role play, as was suggested after the pilot (Eurofound and Cedefop, 2018a). Interviewers considered the training material useful.

During fieldwork, supervisors listened-in on phone calls and/or recorded phone calls of at least 10% of the screener interviews. In addition, a random sample of the closed screeners was re-contacted by a supervisor to verify the quality. Moreover, local project coordinators received performance statistics by interviewer on a regular basis from the central coordination office. These quality checks

allowed to continuously monitor interviewer performance and undertake the necessary steps if interviewers deviated from script. In several countries, supervisors in charge of day-to-day management regularly convened meetings with interviewers to discuss best practices to get past the gatekeeper and to enhance mutual learning.

Table 11: Accuracy indicators related to fieldwork training, scripting, fieldwork monitoring, data validation and data processing

No.	Indicator	Target	Result	Assessment
Fieldwork training				
7.3 (RQ)	Interviewer training materials are comprehensive, fully covering the process for respondent selection and recruitment as well as for questionnaire administration.	Yes	Yes	✓
7.6 (RQ)	Percentage of interviewers that take part in the training	100%	100%	✓
Scripting				
6.1 (RW)	Number of scripting errors encountered in the pilot test	0	1	✗
6.4 (RQ)	Number of consistency rules (soft and hard) identified	>0	Yes	✓
6.5 (RW)	Percentage of identified consistency rules integrated in the script	100%	100%	✓
Fieldwork monitoring				
10.1 (RQ)	An explicit, comprehensive and discriminative interviewer monitoring strategy is outlined	Yes	Yes	✓
10.3 (RW)	Percentage of issues identified based on information in weekly monitoring data for which a solution is provided	100%	88%	✗
10.4 (RW)	Percentage of countries where at least 10% of successful screener interviews are monitored	100%	100%	✓
10.5 (RW)	Percentage of countries where at least 10% of follow-up calls for CAWI completion are monitored	100%	100%	✓
Data validation				
10.10 (RQ)	An explicit, comprehensive and discriminative data validation strategy is outlined	Yes	Yes	✓
10.12 (RW)	The approach to data validation is applied in real time during data collection	Yes	Yes	✓
10.13 (RW)	Percentage of issues identified based on data validation for which a solution is provided	100%	100%	✓
Data processing				
11.1 (RQ)	A systematic approach to coding multi-format items into a single format and for collapsing complex/long questions is in place	Yes	Yes	✓

Source: (Eurofound and Cedefop, 2020b)

Note: RQ: required indicator; RW: real world indicator

Field visits to national partners by EF/CF during fieldwork also helped to improve data quality as feedback from these visits was quickly acted upon.¹¹

The MM and ER questionnaire was programmed using a common script in all countries. This script included routing checks so that each question was only answered if appropriate based on previous responses. The script also conducted permitted values checks so that respondents could only provide sensible answers. On the first day of fieldwork a routing error was detected in Estonia, and quickly corrected (Eurofound and Cedefop, 2020e, p.9). As mentioned earlier, a translation error in the local language script in Sweden was detected during mainstage fieldwork and led to a substantial loss of observations.

In addition, the web survey included hard and soft consistency checks. Both types of consistency checks notify a respondent when an answer is illogical (e.g. the number of managers is larger than the total number of employees; a question is skipped). Respondents must correct the hard checks before they can proceed, whereas soft checks can be overruled. The number of hard and soft checks shown to a respondent was recorded and used to assess data quality. The median number of hard and soft checks shown to the respondent in the MM survey was 1 (Eurofound and Cedefop, 2020e, Table 13). 56% of the respondents were shown at least one warning message.

During data processing, several indicators were used to identify poor data quality (see data quality and cleaning report for details). More specifically, each survey was checked for item non-response; survey length and speeding; straight lining; the number of consistency checks shown to the respondent; and implausible responses (on ranking questions, the hierarchy question, and the skills questions). The number of employees reported was checked for outliers. Based on this information (for details, see EF/CF (2019b), Table 5), an overall quality score per respondent was calculated. Respondents with a low overall quality score were removed from the dataset. In total, 161 MM interviews (0.74% of the total) and 22 ER interviews (0.71% of the total) were dropped. The proportion of MM interviews dropped is fairly similar across countries, and ranges from 0% in Estonia to 1.66% (6 cases) in Slovakia. However, while the Stata syntax to calculate the overall quality score is included in the appendix of the data quality and cleaning report, it is unclear how the threshold has been set to discard observations. **It might be useful to clearly document the quality score threshold in the data quality report. It might also be considered to include the overall quality score in the datasets so that users can decide which cases to exclude from their analyses or can test whether their results are robust to excluding low quality cases.**

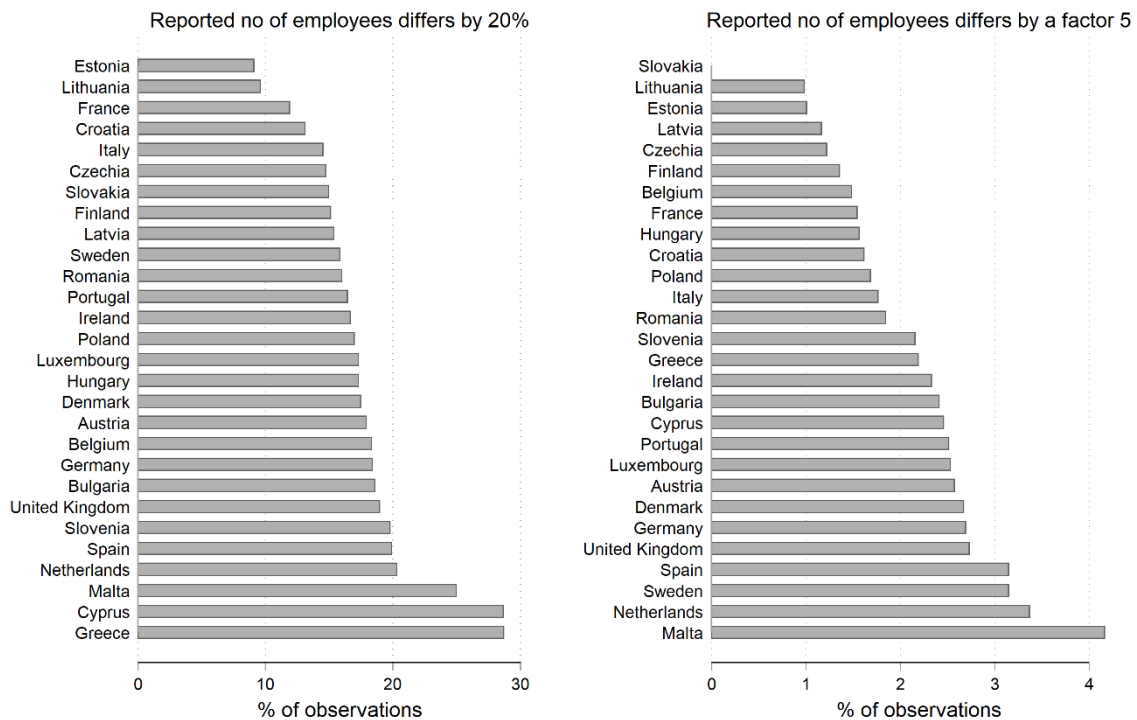
The data quality and cleaning report flagged outliers with regard to the number of reported employees. The most extreme data point was observed in the UK, where one establishment reported employing 160,000 employees. As of now, the choice to include or exclude these outliers when analysing the data is left to the data user. It could be considered to flag those outliers in the final dataset or, even, to recode the most extreme outliers as missing variables. Data users could then decide to work with a 'cleaned' version of the dataset or verify whether their findings hold when using the 'cleaned' dataset.

The size of an establishment is one of the key variables in the ECS. Moreover, it is one of the few questions that is asked in both the screener interview and the MM interview. As mentioned above,

¹¹ In Greece, for instance, the field visits revealed that establishments were re-contacted for quality purposes during the period in which reminders were sent, generating confusion for respondents. This was rectified and quality assurance calls were scheduled after the final reminder.

the quality and data cleaning report identifies outliers with respect to the number of employees as reported by the manager. The responses are, however, not compared between the screener interview and MM interview. Such a comparison provides an additional quality control. In 18% of the cases, the number of employees reported in the screener and MM interview differs by at least 20%, with substantial variation across countries (Figure 10). Both estimates even differ by a factor of five in 3% of the observations. **More detailed analyses, preferably in real time during the data collection so that mistakes can be corrected on the spot, could improve data quality in future waves.** For instance, an additional warning message could be included in the online survey which warns respondents when the reported number of employees differs substantially from the number reported during the screener interview or from the number in the sampling frame. Prompted by the warning message, respondents will then correct their response if they made a mistake (for instance, correcting typing errors, such as reporting 160,000 employees, while the real number is 16,000).

Figure 10: Comparing the number of employees reported in the screener interview versus the MM interview



Source: author calculations on ECS 2019 management dataset using the variables 'wpsize_n' (establishment size reported in the screener) and 'wpsize_mm' (establishment size reported in the MM interview)

2.4. Weighting

An important element of a survey is its representativeness for the population under study. Weights are used to correct for disproportionalities between the sample and the population due to the sampling strategy and non-response bias. Weights in the ECS were developed in five steps which correspond to the different steps in the survey process (Eurofound and Cedefop, 2020d):

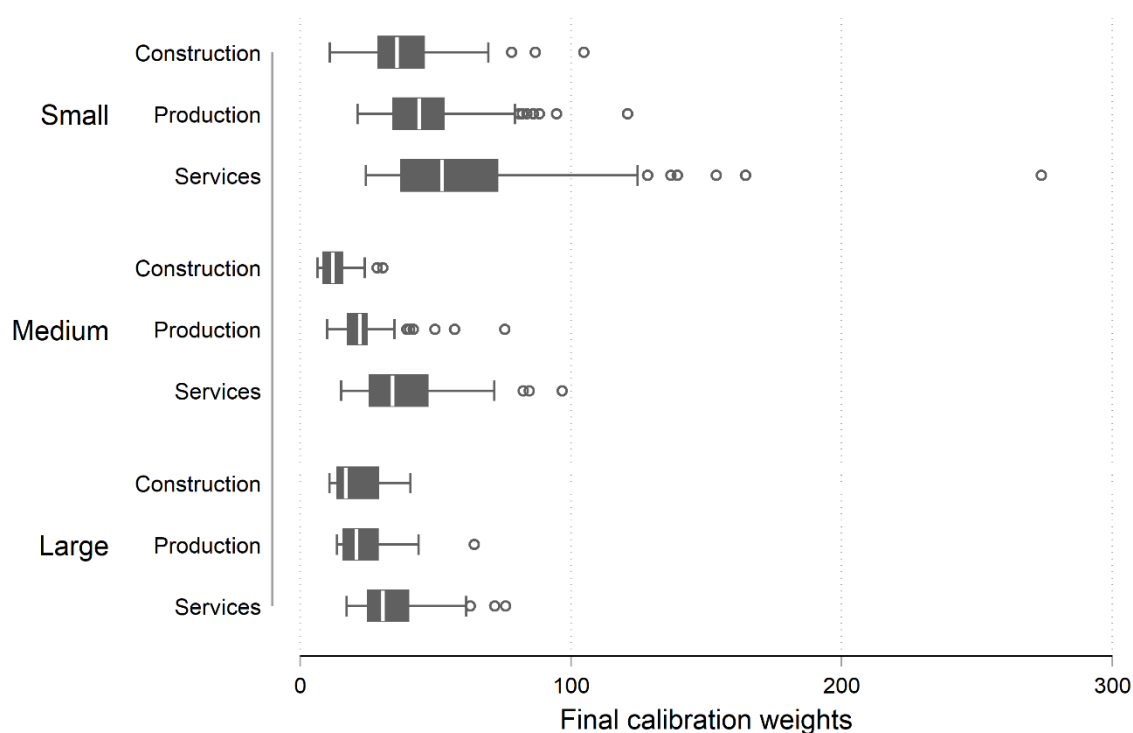
1. *Design weights*, which correct for disproportionalities due to the sampling design;
2. *Non-response adjustment (screener stage)*, which takes into account that participation in the screener interview may depend on the sector (top level NACE), establishment size (small, medium, or large) and region (NUTS 3);

3. *Non-response adjustment (CAWI interview stage)*, which takes into account that participation to the online survey (after a successful screener interview) may depend on the sector (top level NACE), establishment size (small, medium, or large) and region (NUTS 3);
4. *Design weights for establishment selection in company-level sampling frame countries*, which take into account that in countries where a company-level sampling frame was used, establishments from multi-site companies were randomly selected;
5. *Calibration weights*, which adjusted the weights so that key population estimates (sector and establishment size)¹² were matched.

In each step weights were trimmed to avoid very small or very large weights. The ‘trimming rules’ differed in each step. Details can be found in the sampling and weighting report.

Due to the weighting procedures weights differ across establishments within a same cell (establishments with the same size and operating in the same sector), as illustrated in Figure 11.

Figure 11: Distribution of weights in Belgium, by establishment size and sector



Source: author calculations on ECS 2019 management dataset (variable: s5_wgt_final)

It shows the distribution of weights for establishments in Belgium within the nine cells. Weights are generally smaller in the construction sector and larger in the service sector. The reason is that the sample in Belgium includes relatively many (few) establishments in the construction (service) sector relative to their real share in the population.¹³ Given a sector, weights also tend to be smaller in

¹² Given the small sample sizes in many of the countries, the same grouped categories were used as those used as sampling strata: the establishment size grouped in three categories (small, medium and large) and the sectors grouped by construction (NACE F), production (NACE B-E) and services (NACE G-N, R and S).

¹³ This does not imply that the sample includes many establishments in the construction sector in absolute terms. For instance, in Belgium, the sample only includes 7 large establishments in the construction sector versus 493 small establishments in the service sector. Relative to the population, however, the large

larger establishments compared to medium and small establishments. This is due to the sampling procedure which oversampled larger establishments (cell quotas were set as the mid-point of establishment-level and employee-level estimates) (Eurofound and Cedefop, 2020d, p.14). The sampling design corrects for this disproportionality between the sample and the distribution by applying appropriate weights. But even within a cell, a substantial variation in weights is observed. For instance, weights of small establishments in the construction sector range from 10.9 to 104.7.

The sampling and weighting report shows that the weighted distribution approximates the population distribution of the establishments by size and sector (Eurofound and Cedefop, 2020d, Table 5). This indicates that weighting indeed improves the representativeness of the survey. Six out of seven QA indicators related to weighting were met (Table 12). Indicator 2.13 was not met because in Greece and Luxembourg no recent reference statistics were available.

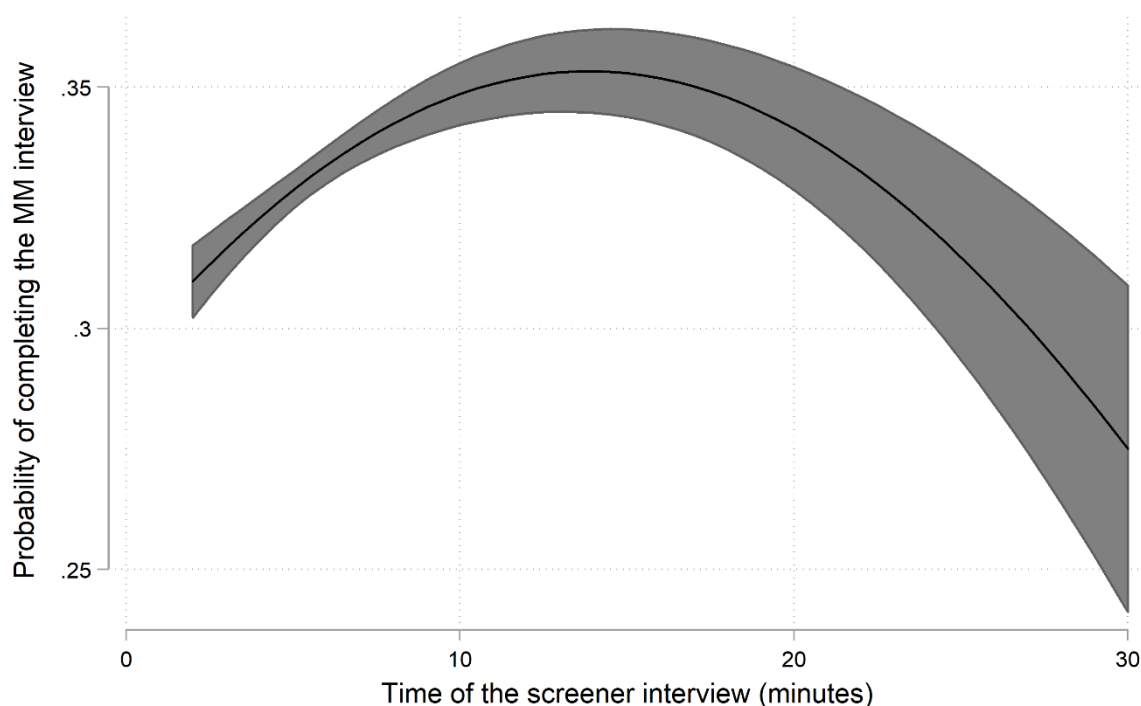
A high level of non-response in social science surveys is common and, as discussed in a previous section, the ECS is no exception. As mentioned above, the second and third step of the weighting procedure accounts for non-response at the screening and CAWI interview stage, taking into account the sector, establishment size and region. This adjustment assumes that the missing data is missing at random (MAR), implying that the probability of response does not depend on the variable of interest (the 'y-variable') once we control for a vector of x-variables from the sampling frame (Groves, 2006; National Research Council, 2013). The survey methodology literature has highlighted the potential of paradata and, more specifically, level-of-effort paradata (e.g. number of times an establishment has been contacted; timing of the contact) to test the MAR assumption and to correct for non-response bias (Biemer, Chen, & Wang, 2013). While it goes beyond the scope of this paper to assess the potential of such an approach, we do believe that it is worth exploring. To illustrate how this could be done in practice, we show that the length of the screener interview is correlated with the probability of completing the online MM survey (Figure 12).

There exists an inverse U-shaped relation between the length of the screener interview and the probability of completing the MM survey. This association could be exploited to calibrate the weights and to test the MAR assumption. Similar level-of-effort paradata could potentially be exploited to correct for non-response at the CATI stage. For instance, the number of times an establishment has been contacted, the timing of the contact and an indicator of the 'persistence' of the interviewer could be used to model the propensity to respond. Whether this approach would really reduce non-response bias¹⁴ is hard to tell, but given that the ECS already collects detailed paradata, it might be worthwhile to commission a feasibility study that examines the potential of such an approach. Such a feasibility study could exploit the ECS 2019 data and the recommendations could then be incorporated in the next ECS wave.

establishments in the construction sector are still oversampled relative to the small establishments in the service sector.

¹⁴ Note that it cannot be tested whether the ECS suffers from non-response bias. A low response rate does not necessarily lead to non-response bias. This is only the case if establishments that do not respond differ systematically from those that do respond (Rindfuss, 2015).

Figure 12: Probability of completing the MM web survey in function of the length of the screener interview



Note: Results from a logistic regression that controls for firm size (small, medium, large), sector (production, construction, services) and country. Full results are reported in the Table 23 in the Appendix 1.

The non-response analysis could also inform sampling design in future waves and determine the initial size of the gross issued sampled. For instance, Ipsos argues convincingly that the construction sector (Nace F) needs to be included as a separate stratum because a previous survey had shown that the yield rate tends to be lower in the construction sector.

A similar analysis of the yield rate based on the ECS 2019 could help to decide how sectors could be grouped so that sectors with substantially different response rates could be included as separate strata. Similarly, an analysis of the yield rate by establishment size and country, could help to inform the sample size that needs to be issued to reach the targets and to minimize the sample that remains 'live' at the end of the survey. For instance, large firms in Belgium are less likely to participate in the screener interview¹⁵, so the initial sample should include more large firms to reach the targets.

¹⁵ See the logit model explaining successful CATI interviews in Belgium (ECS4_CATI_model_log_BE).

Table 12: Accuracy indicators related to weighting

No.	Indicator	Target	Results	Assessment
Weighting strategy				
2.1 (RQ)	A common weighting strategy is agreed that reflects current academic standards, as illustrated by academic references included in the weighting strategy document, and integrates all available information on those elements that are foreseen to be included in the weighting procedure, given the sampling plan	Yes	Yes	✓
2.3 (RW)	Percentage of countries where the agreed weighting strategy is used	100%	100%	✓
Design weight				
2.4 (RQ)	Percentage of countries where the design weight is specified in accordance with the sampling design	100%	100%	✓
Post-stratification weight				
2.8 (RQ)	Percentage of countries where the post-stratification weight takes all agreed variables into account	100%	100%	✓
2.9 (RW)	Percentage of countries where the same agreed set of variables with the agreed categories are used for weighting	100%	100%	✓
Supra-national weights				
2.13 (RW)	Percentage of countries where the weights are based on up-to-date official population statistics collected within two years preceding fieldwork	100%	93%	✗
Size of weights				
2.17 (RQ)	The size of the weights is limited (e.g. through trimming or weight shrinkage), based on a strategy that strikes an appropriate balance between representativeness in terms of the weighting variables and the effective samples size, which is fully documented and replicable	Yes	Yes	✓

Source: (Eurofound and Cedefop, 2020b)

Note: RQ: required indicator; RW: real world indicator

2.5. Data and results dissemination

The ECS is unique in that it is the only EU-wide company survey about workplace practices surveying establishments. It addresses policy-relevant topics. This makes the survey highly relevant for researchers as well as policymakers. The data is well documented and easily accessible currently through the UK Data Service. The four QA indicators related to the accuracy of the micro-data were all met (Table 14).

Yet, the data appear to be somewhat underused in the academic community. There is room to enhance the usage and visibility of the data among the research and policy communities. We support the recommendations of the feasibility report 'Methodological lessons learnt from the third European Company Survey' to promote the use of the data and suggest to, if possible, continue implementing these recommendations. The Cedefop, Eurofound and IZA Conference on Workplace and Management Practices, which will be organised in August 2020 in Dublin, the special issue of the

Disclaimer: This working paper has not been subject to the full Cedefop/Eurofound evaluation, editorial and publication process.

academic journal *Research in Labor Economics* on workplace practices and the ECS 2019 overview report launch event, which will take place in Brussels in October 2020, are already excellent initiatives to promote and disseminate the data.

The feasibility report mentioned above analysed the use of the data of the different ECS waves until 2016. We extend this analysis to October 2019 using information on the number of times the data was downloaded from the UK Data Service (information provided by Eurofound on October 29, 2019).

Table 13 summarises characteristics of the users. From June 2016 until October 2019, the ECS 2013 data was downloaded 462 times. We observe a peak in 2016 and 2017 when the data was downloaded over 100 times. Download numbers were fairly similar for previous waves (Eurofound, 2017a, Figure 5). Across all years, the data have been downloaded in 36 countries, but users are predominantly based in the UK (137 downloads), followed by Spain (46), Italy (43), the Netherlands (40), Germany (37) and Belgium (23). In 12 out of the 28 countries included in the ECS, the data were downloaded less than five times. The data is almost exclusively used for research. Only three users downloaded the data for teaching purposes.

Table 13: Downloads of the ECS 2013 data from the UK Data Service

	Times downloaded	%
Times downloaded	462	100%
By year:		
2016	123	27%
2017	111	24%
2018	72	16%
2019	62	13%
By country:		
United Kingdom	137	30%
Spain	46	10%
Italy	43	9%
Netherlands	40	9%
Germany	37	8%
Belgium	23	5%
Other countries	117	25%
Type of use:		
Teaching	3	1%
Non-commercial	458	99%

Source: UK Data Service

Note: Data obtained on October 29, 2019. The country and type of use was not reported by respectively 19 and 1 users

Monitoring downloads from the UK Data Service is one way to assess the use and impact of the dataset. But, the ECS 2013 also reaches stakeholders through other channels, most importantly through the publication of the overview reports. Google scholar indicates that the overview report

‘Third European Company survey – Workplace innovation in European companies’ was cited 33 times as of April 2020. Overview reports using the ESENER-2 dataset and the EWCS 2020 dataset were cited respectively 15 and 49 times.

We suggest continuing monitoring the number of downloads as well as the number of papers and reports using ECS data or making reference to publications from EF/CF and other sources on the ECS, so that the impact of dissemination activities related to the release of the ECS 2019 data is closely monitored. In addition to the dissemination activities mentioned earlier, it might be useful to set up additional dissemination activities to actively promote the data. One could think of facilitating publications in special issues, providing grants to researchers analysing the data, awarding a best paper award to a paper using ECS data or giving researchers (through an open competition) the opportunity to propose questions to be included in an ad-hoc module in a future wave under the condition that the responses are analysed and the results published.

Table 14: Accuracy indicators related to the micro-data

No.	Indicator	Target	Results	Assessment
12.2 (RQ)	Percentage of variables in the dataset that are named and labelled in accordance with the agreed template	100%	100%	✓
12.3 (RQ)	Percentage of variables for which the metadata (e.g. missing values, measurement level) are properly defined in the dataset	100%	100%	✓
12.4 (RQ)	Percentage of substantive variables included in the dataset	100%	100%	✓
12.5 (RQ)	Percentage of auxiliary variables (contact data, paradata, frame data etc.) included in the dataset	100%	100%	✓

Source: (Eurofound and Cedefop, 2020b)

Note: RQ: required indicator; RW: real world indicator

2.6. The Quality Assurance Plan

A Quality Assurance Plan (QAP), initially developed by Eurofound, was used to monitor survey processes and ensure high-quality data. The QAP consisted of 129 indicators¹⁶, structured by seven survey stages and five quality dimensions (Table 15). These indicators were updated monthly by Ipsos and discussed with EF/CF. A distinction was made between ‘required’, ‘real world’ and ‘ideal world’ targets.

In its quality control report, IPSOS stated that *“The QC [Quality Control] approach adopted for the 2019 ECS was a reasonably effective and an efficient process. The volume of indicators in total was not ideal but given that not all indicators applied at all stages of the survey this could be managed successfully to avoid over-burden in assessment and reporting.”* (Eurofound and Cedefop, 2020b, p.65). In addition to this general assessment, it offers a few specific recommendations to improve the QAP, including adding two indicators to avoid errors in the translations (as occurred in Sweden) and in the sample delivery (as occurred in Slovenia).

¹⁶ The initial QAP contained 132 indicators, but three indicators were dropped during implementation. Out of the 129 remaining indicators, 3 indicators were ‘not applicable’ or remained ‘to be confirmed’ by the end of the project

Table 15: Overview of quality assurance indicators by quality dimension

	Accessibility	Accuracy	Coherence & comparability	Punctuality	Relevance & timeliness	TOTAL
Sampling	4	13	6	6		29
Weighting	6	7		4		17
Questionnaire	3	5	1	4	1	14
Translation	5		5	5		15
Pilot		5				5
Fieldwork infrastructure			1			1
Scripting	1	3	1	1		6
Fieldwork training		2	2	3		7
Mainstage fieldwork and data validation	2	14		6		22
Data processing	1	1	1			3
Micro data	2	4		2		8
Reporting	1			1		2
TOTAL	25	54	17	32	1	129

Source: (Eurofound and Cedefop, 2020c)

In our view, several elements have contributed to the successful adoption of the QAP in the ECS 2019. First, several discussions prior to fieldwork between EF/CF and Ipsos helped to clarify the indicators' objectives. A few indicators were also revised. Moreover, the QAP was a live document and indicators could still be revised throughout the survey's implementation. Second, a designated Delivery Manager at Ipsos was responsible for updating the QA indicators. This helped because she was very familiar with the QAP. Third, monthly meetings between EF/CF and Ipsos to discuss the QA indicators ensured that the QA indicators were closely monitored and informed fieldwork. These practices should also be adopted in future waves.

Nevertheless, in our view a few indicators could be revised so that the outcome is measurable. In particular, indicators such as 1.1. (*"A common sampling strategy is agreed that reflects current academic standards, as illustrated by academic references included in the sampling strategy document"*) or 3.7 (*"Clear translation instructions"*) are not clearly defined.

We believe that the QAP is a good instrument to monitor the different survey stages. It helps to quickly identify important quality issues and to quickly resolve them. At the same time, relevant QA indicators cannot capture all issues that can go wrong in a complex survey. Examples of issues that went 'wrong' in the ECS 2019 and were not captured by the QA indicators include the translation error in Sweden or the incomplete sample in Slovenia, but there are many other issues that could have gone wrong and would not have been captured by the QA indicators. Moreover, it is difficult to set targets for some important quality issues, because the targets would be really hard to reach or

because it is hard to tell if the targets are feasible before the start of the survey. For instance, one of the main limitations of the ECS 2019 is the low number of achieved ER interviews, but this is not apparent from the QA indicators. There were, however, good reasons not to set specific targets for the number of ER interviews. Hence, the QAP is only one element to monitor and ensure high-quality data.

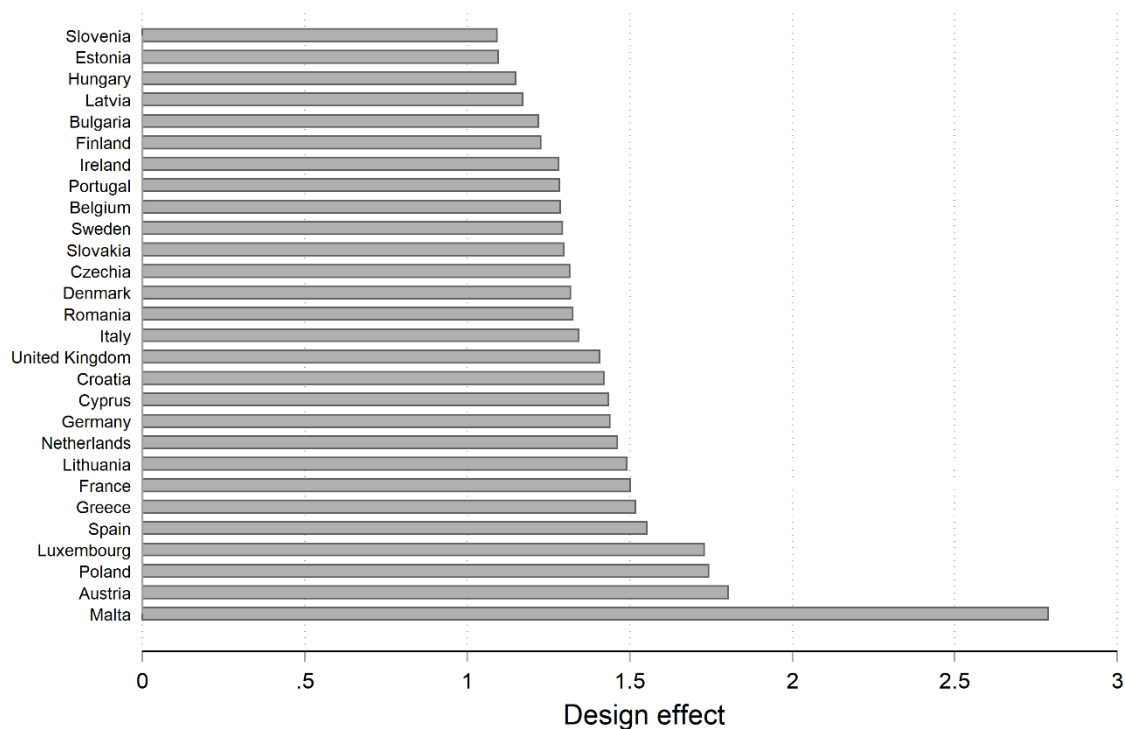
3. Assessment of data quality

3.1. Internal validity

Internal validity refers to the extent to which the concepts are measured accurately and precisely. Accurate measurement implies that the concept that one intends to measure is indeed being captured. Measurement errors induced by, for instance, respondents who did not correctly understand the question or straight-lined questions, cause inaccurate measurement. Assessing measurement error is challenging because there is no ‘gold standard’ against which the response can be compared. Nevertheless, as discussed above, the survey followed best practice to minimise measurement error. The section on ‘monitoring data quality’ offers a few specific suggestions for additional, real-time data quality checks.

The precision of a measurement is easier to assess as it relates to the sampling design. One common indicator of the precision of a survey is the so-called design effect, which is reported in the sampling and weighting report. The design effect indicates the adjustment made to the sample size due to the sampling design compared to simple random sampling. The larger the design effect, the lower the effective sample size. Figure 13 shows the design effect by country. The design effect in, for instance, Belgium is 1.29, which means that the effective sample size in Belgium is 715 ($1,011/1.29$). With the exception of Malta, the design effect is lower than 2 in each country, which is similar as the design effect in other EU-wide surveys. The design effects in the 4th EQLS, for instance, range from 1.13 to 2.30 (EF, 2018, Table 3). The sampling and weighting report states that the large design effect in Malta is due to substantial differences between the unweighted sample and the population distribution, the small sample size achieved and the high proportion of stratum jumpers (establishments misclassified in the sampling frame).

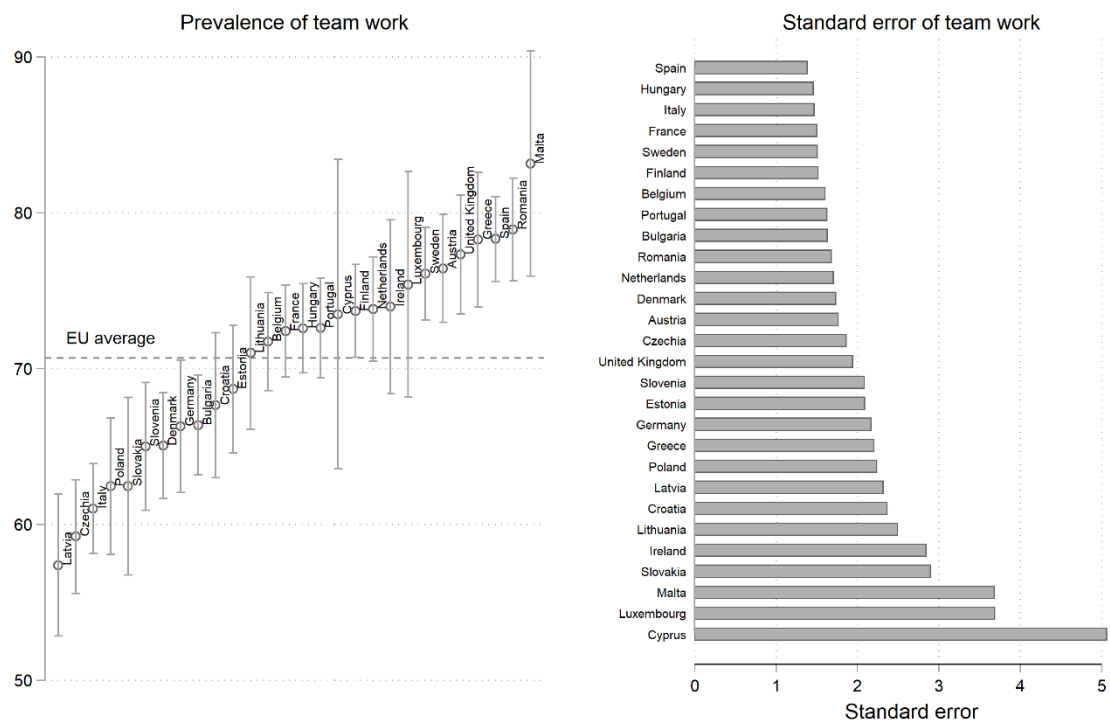
Figure 13: Design effect by country



Source: Own compilation based on Table 23 (Eurofound and Cedefop, 2020d)

A second standard approach to assess precision are the standard errors. Standard errors depend on the sample size and sampling design as well as the specific variable that is being considered. To illustrate this approach, we calculated the standard errors and 95% confidence intervals of the prevalence of team work across establishments in the 28 EU Member States (Figure 14), taking into account the weights and the design of the survey. While the differences between EU countries are interesting, we are primarily interested in the precision of the estimates. The smallest standard error (1.39%) is observed in Spain; the largest in Cyprus (5.06%). Hence, the 95% confidence interval in Spain ranges from 75.6% to 81.0%, whereas the 95% confidence interval of Cyprus ranges from 63.6% to 83.4%. The standard errors and confidence intervals are thus relatively large, particularly when the sample size is low such as in Cyprus, Luxembourg and Malta. **Although it is tempting to use the ECS 2019 to rank and compare countries, the large standard errors imply that such a comparison warrants caution**, as even relatively large differences between countries might not be statistically significant. At the same time, EU-wide surveys like the ECS are often used to compile statistics at the EU-level, by welfare regime or by geographic clusters (e.g. the Baltic States or southern Europe). Such an aggregation increases the number of observations and, hence, the precision of the estimates.

Figure 14: Teamwork across the EU: precision of the estimates



Source: author calculations on ECS 2019 management dataset (variable: 'teamex')

Note: Estimates are weighted (variable: 's5_wgt_final') and standard errors are corrected for the survey design

Measurement error can depend on the respondent's characteristics (Gaziano, 2005). For instance, older managers may have a different view on or a different understanding of certain concepts than younger managers. Hence, one could get very different results for the same establishment depending on the manager that is interviewed. We cannot evaluate measurement error induced by respondents' characteristics. This is always difficult, but even more so in the case of the ECS because there is almost no information about the respondents, with the exception of position (general manager, owner-manager ...) and sex. **It would be useful to collect at least a minimum of**

information about the respondent (age, educational level, sex, number of years employed by the company ...). Including many new items in the survey is expensive and may not be feasible. At the same time, it is hard to tell which respondent characteristics really affect measurement error. As an alternative, one could consider splitting the sample in subsamples and ask different questions to each subsample. By comparing responses within and between subsamples, one could assess which respondent characteristic determine measurement error. Such an experimental approach could also be useful to test and validate more substantive questions such as those measuring teamwork or job autonomy.

3.2. External validity

In this section, we attempt to assess the external validity of the ECS. External validity refers to the extent to which findings from the ECS are similar to findings from other surveys or administrative datasets that measure the same concepts. Ideally, the ECS findings should be compared to the ‘true’ value in the population. Unfortunately, few national or international surveys exist that measure the same concepts for the same population, while administrative datasets do not exist or are not easily accessible. Moreover, even when such datasets exist, they cannot claim to measure the ‘true’ value either. In other words, there exists no gold standard or benchmark against which the quality of the ECS can be assessed.

As there is no gold standard, we rely on a different strategy to assess the external validity of the ECS. This strategy helps to identify poorly measured variables and helps to identify countries where the ECS might not accurately capture working conditions. It gives, however, only an indication of external validity.

The strategy consists of two steps. We first compare key indicators from the ECS 2019 to the same indicators from the ECS 2013. This allows us to study trends over time, which is already interesting in itself. However, it is unclear whether these differences are real, that is caused by real (institutional) changes in a country between 2013 and 2019, or spurious, that is caused by data quality issues. In order to disentangle real from spurious changes, trends observed in the ECS are compared to trends observed in other datasets measuring similar, correlated concepts. For instance, if according to the ECS, the presence of ER in establishments decreased from 2013 to 2019, one would also expect a (commensurate) decrease in trade union density as measured by the OECD.¹⁷ This strategy can be expressed as follows:

$$\nabla y_{it,ECS} = \beta \nabla y_{it,d} + \delta \quad (1)$$

We will estimate whether the change over time of a particular indicator in country i as measured by the ECS ($\nabla y_{it,ECS}$) can be explained by the change in a similar indicator from a different dataset ($\nabla y_{it,d}$). The parameter δ then indicates to what extent findings from the ECS differ from findings from other datasets. If δ is not statistically different from zero, then the change observed by the ECS can be explained by the change in the indicator from a different dataset. This suggests external validity of the observed change. In other words, it would indicate that both measurements point towards similar trends. On the other hand, if δ is statistically different from zero, then the ECS (or

¹⁷ An alternative dataset that contains more detailed information is the ICTWSS dataset, though data for 2019 are often not yet available (<http://uva-aias.net/en/ictwss>).

the other dataset) is either biased or changes in the indicator from the external dataset do not (fully) explain changes in the ECS indicator.

This approach is illustrated for two indicators: (1) the presence of an ER and (2) the share of employees participating in training during paid working time.

This strategy could be repeated for other ECS indicators for which correlated indicators from international datasets are publicly available such as indicators related to (skills) mismatch or the prevalence of open-ended contracts.

Is the presence of an ER in the establishment accurately measured?

The presence of an ER in the establishment is one of the key indicators of the ECS. The screener interview already asks the respondent to identify an ER who then receives the ER questionnaire.¹⁸ In addition, the management questionnaire includes the question “*Which of the following forms of official employee representation currently exist in this establishment?*”. The ECS 2013 MM questionnaire included a similar question.¹⁹

Using these data, the proportion of establishments with an ER in 2013 and 2019 was estimated (Figure 15). In order to make the population as comparable as possible, establishments in the public sector were excluded from the ECS 2013. The proportion of establishments with an ER differs considerably across countries. More importantly, this proportion decreased in nearly all countries from 2013 to 2019 and, often, substantially so. In Denmark and the Netherlands, for instance, the prevalence of ER decreased by approximately 25 percentage points.²⁰

We use another indicator ‘trade union density’ to verify whether these trends are ‘real’ or caused by data quality issues. The indicator ‘trade union density’ is compiled by the OECD and available for 20 and 15 EU Member States in 2013 and 2018 respectively.²¹ As discussed earlier, the assumption is that changes in trade union density are correlated to the presence of an ER.²² In both years, a positive correlation between trade union density and the share of ER in establishments is observed (Figure 19). The correlation is respectively 0.54 in 2013 and 0.50 in 2019.

Trade union density has decreased from 2013 to 2018. This decrease is, however, too small to explain the decreasing share of ER. Table 16 estimates regression 1. δ equals -9.41, which implies that – even when accounting for decreasing trade union density – the unexplained difference between the share of ER in 2019 and 2013 equals 9.41 percentage points.

¹⁸ Without taking the sampling weights into account, the proportion of establishments with an ER across the EU is slightly lower in the screener dataset (35%) than in the MM dataset (37%). However, the difference between both datasets is substantial for a few countries. For instance, in the UK, 31% of the establishments have an ER according to the screener dataset, compared to 24% according to the MM dataset.

¹⁹ The different forms of ER were not identical in the ECS 2013 and the ECS 2019. The ECS 2013 identified 7 generic ER types; the ECS 2019 identified 8 types. This difference may contribute to explaining trends in the prevalence of ER from 2013 to 2019.

²⁰ In the Netherlands, both waves identified two ER types (the ‘Works Council’ and ‘the public equivalent of a work council’). By contrast, the ECS 2019 identified two ER types in Denmark, whereas the ECS 2013 identified three types.

²¹ Data for 2019 are not yet available.

²² <https://stats.oecd.org/Index.aspx?DataSetCode=TUD>

Table 16: Explaining the change in the presence of an ER between 2013 and 2019

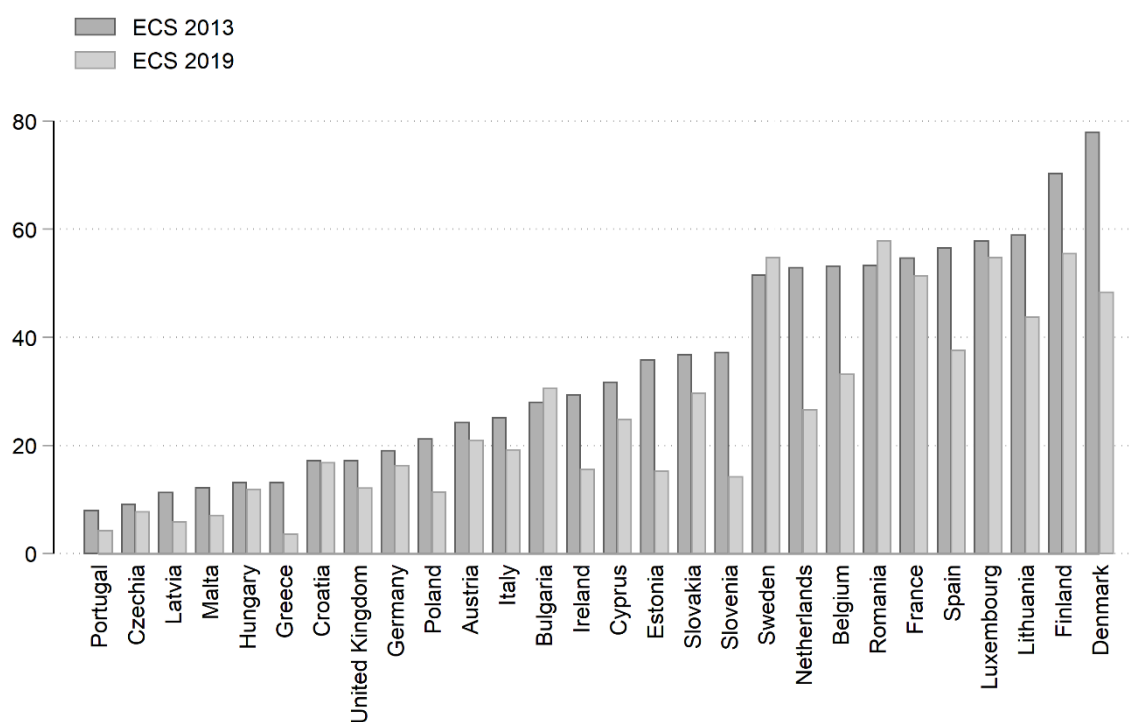
Variable	Coefficient
β	0.817
	(1.374)
δ	-9.410**
	(4.304)
N	15
R-squared	0.027

Source: author calculations on ECS 2019 management dataset

Note: Standard errors in parentheses. ***, **, * denote statistical significance at the 1%, 5% and 10% level, respectively

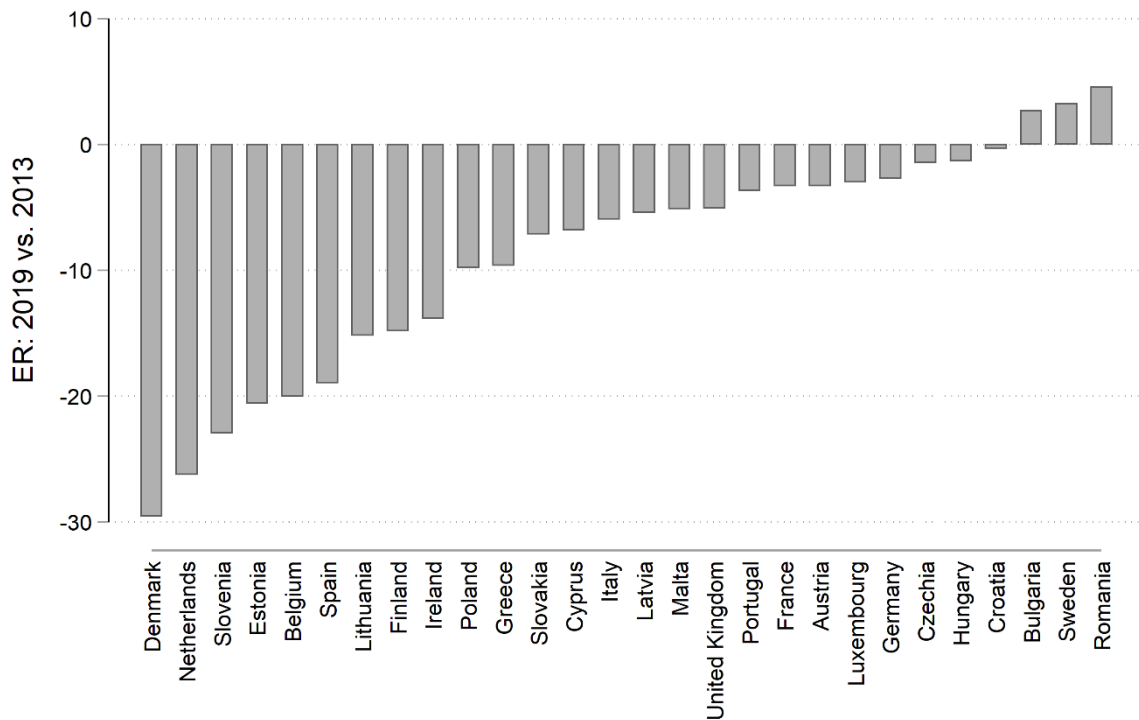
This unexplained difference is rather substantial. It might be related to data quality issues such as the switch to the push-to-web methodology and the difficulties encountered in identifying ER in the screener interview. We can, however, not rule out that ‘real’ changes explain a part of the unexplained difference. One could, for instance, argue that a declining ER presence in establishments does not necessarily imply a declining trade union density. **More detailed analyses (e.g. the evolution by type of ER and country) would help to explain what is driving the overall decrease in the share of ER.** In addition, it could be useful to discuss with country experts whether they can offer explanations for the significant decrease in the presence of ER from 2013 to 2019 observed in, for instance, the Netherlands or Denmark.

Figure 15: Share of establishments with an ER in 2013 and 2019



Source: author calculations on ECS 2019 management dataset (variables: mmerconfirm_v1-8) & the ECS 2013 MM dataset (excluding establishments in the public sector) (variables: ERTYPE_A – G)

Figure 16: Change in share of establishments with an ER from 2013 to 2019



Source: author calculations on ECS 2019 management dataset (variables: mmerconfirm_v1-8) & the ECS 2013 MM dataset (excluding establishments in the public sector) (variables: ERTYPE_A – G)

How accurately is training during paid working time measured?

Relying on the same strategy, we also verify the external validity of the question ‘In 2018, how many employees in this establishment participated in training sessions on the establishment premises or at other locations during paid working time?’. A similar question was also included in the ECS 2013. Both questions can be compared to Eurostat’s indicator on ‘lifelong learning’. This indicator is derived from the Labour Force Survey and measures the proportion of adults aged 25 to 64 years that reported having participated in formal and non-formal education in the past four weeks. Participation in education and training is arguably a similar concept as and should be correlated to participation in training sessions during paid working time.

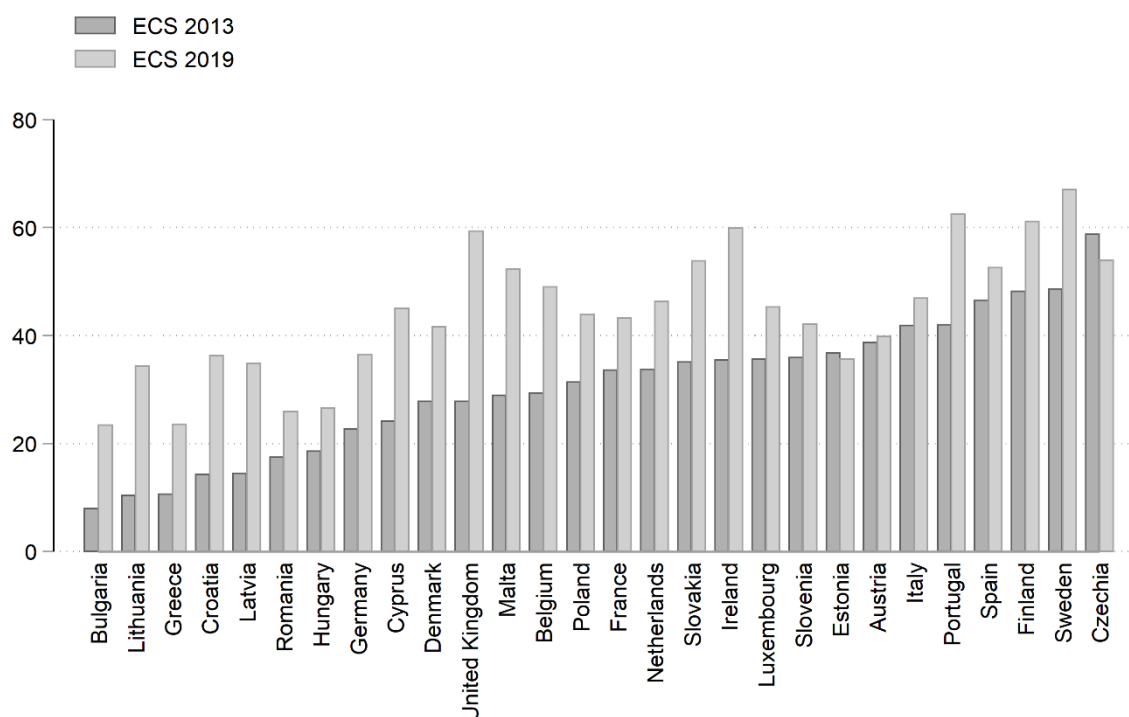
Figure 17 compares the share²³ of employees who participated in a training session during paid working time according to the ECS 2013 and the ECS 2019. Figure 18 evaluates how participation in training evolved from 2013 to 2019. With the exception of Czechia and Estonia, the share of employees participating in training increased in all countries. Participation increased by more than 10 percentage points in 18 out of the 28 countries and even increased by 30 percentage points in the UK.

The ECS indicator is compared to Eurostat’s indicator on lifelong learning. In both waves, there is a positive correlation between participation in training sessions as measured by the ECS and lifelong learning (Figure 20 in Appendix 1). The correlation is respectively 0.54 in 2013 and 0.51 in 2019.

²³ We do not know the exact shares. The share of employees participating in training is reported in percentage categories (0%; <20%; 20-39%; 40%-59%; 60%-79%; 80%-99%; 100%). We use the ‘midpoint’ in each category when estimating the average share in a country.

Hence, countries that excel in lifelong learning also do well with regard to participation in training during paid working time.

Figure 17: Share of employees participating in training sessions during paid working hours



Source: author calculations on ECS 2019 management dataset (variables: paidtrain_d) and ECS 2013 management dataset, excluding establishments in the public sector (variables: HTRAIN)

To verify the external validity, we examine whether changes in Eurostat's indicator explain changes in the ECS indicator. This is not the case. The correlation between changes in both indicators is negative and not significant (Table 17). Moreover, the parameter δ equals 14.05, indicating that, even after controlling for the change in lifelong learning from 2013 to 2018, participation in training has increased by 14.05 percentage points between 2013 and 2019 according to the ECS.

Table 17: Explaining the change in the share of employees participating in training during paid working time

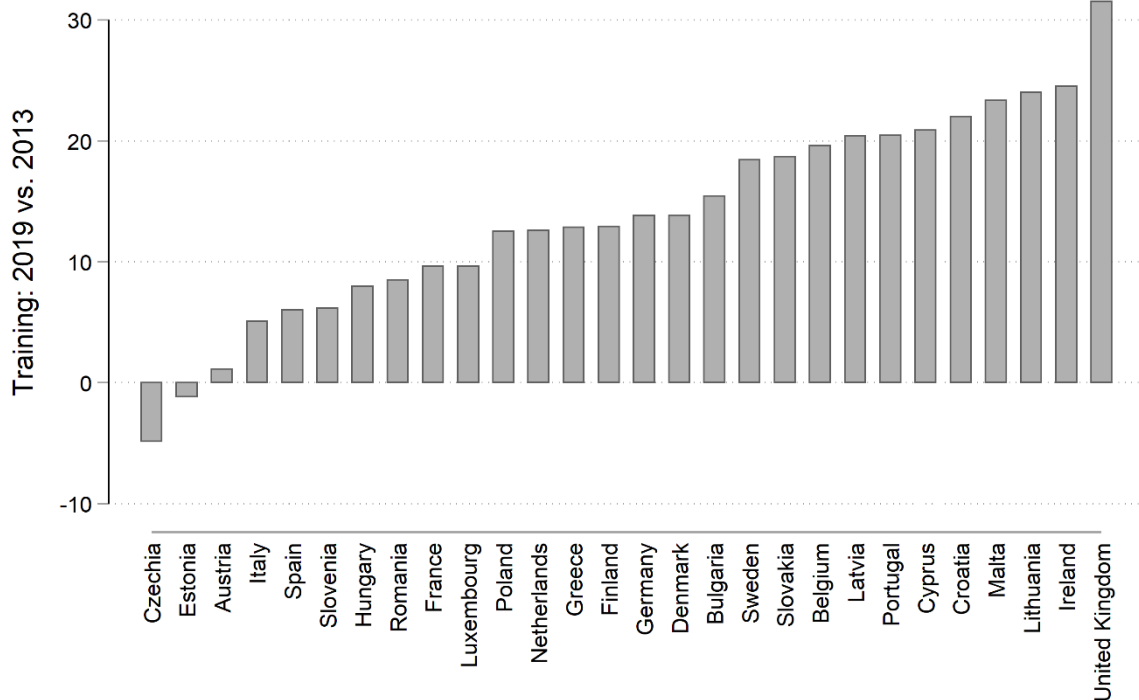
Variable	Coefficient
B	-0.306
	(0.630)
δ	14.05***
	(1.705)
N	28
R-squared	0.009

Source: author calculations on ECS 2019 management dataset

Note: Standard errors in parentheses. ***, **, * denote statistical significance at the 1%, 5% and 10% level, respectively

This finding raises questions about the external validity of the finding that participation in training has increased between 2013 and 2019. Although we cannot rule out that participation in training during working time has indeed increased substantially between 2013 and 2019, it seems likely that methodological changes between the ECS 2013 and ECS 2019 drive the results. The substantial (and, possibly, unrealistic) increase in countries like the UK and Ireland calls for further discussions with country experts who might be able to offer compelling explanations and to assess whether the observed trends are sensible.

Figure 18: Change in participation in training between 2013 and 2019 according to the ECS



Source: author calculations on ECS 2019 management dataset (variables: *paidtrain_d*) and ECS 2013 management dataset, excluding establishments in the public sector (variables: *HTRAIN*)

4. Recommendations

We conclude the quality assessment of the ECS 2019 with several recommendations (Table 18). These recommendations are structured along the survey's lifecycle (questionnaire development, sampling, fieldwork, monitoring data quality, weighting, data and results dissemination). Recommendations related to the external validity and the QAP are then discussed. Most of the recommendations were already briefly mentioned before and are further developed here.

The recommendations are ranked according to their impact on data quality and feasibility. Recommendations with a high, medium and low expected impact are respectively highlighted in red, orange and yellow in Table 18. The feasibility of a recommendation is related to the efforts required to implement it. Recommendations that are difficult (+++), moderately difficult (++) or relatively easy (+) to implement are distinguished.

Two issues are in our view critical to address before implementing the next survey wave: (1) the low number of achieved ER interviews (recommendation 9) and **(2) the relatively limited use of the data** (recommendation 13). The other recommendations could help to further improve the survey's quality, but are considered less of a priority. Each recommendation is discussed in detail below.

1. [Questionnaire development] Develop a research plan for each (new) question (*medium impact; moderately difficult to implement*)

Each (new) research item/question introduced in the ECS should be accompanied with a (detailed) research plan. This plan should indicate which policy-relevant issue the question addresses, how the question relates to the other questions in the questionnaire and how the question will be analysed. Such a research plan helps to avoid that too many questions are included (which complicates the questionnaire and leads to an additional burden for the respondents) and could contribute to ensuring that all questions are used to inform policies.

2. [Questionnaire development] Determine a set of core questions that are asked in each wave so that time trends can be identified (*medium impact; moderately difficult to implement*)

The ECS does not aim to identify time trends. The questionnaire has changed in each wave. Moreover, the shift to a push-to-web methodology implies that the findings of the ECS 2019 cannot easily be compared to the findings of the ECS 2013 or those of earlier waves. However, it would be valuable to determine a set of core questions that are asked in each wave, even if the survey mode changes. This would increase the survey's impact (particularly in the longer term) as it would allow data users to study, at least to some extent, time trends. In addition, studying time trends helps assessing the external validity of the ECS (see recommendation 15). If a core set of questions is selected, it is important to keep the wording and, if possible, the order of these questions unaltered between waves so that responses can be compared. It is also recommended to use the same coding scheme across different waves so that the different waves can relatively easily be linked or to make a dataset publicly available that already includes all waves.

Table 18: Recommendations

Recommendation Colour code: High impact: red; medium impact: orange; low impact: yellow	Feasibility Difficult: +++ Moderately difficult: ++ Relatively easy: +
Questionnaire development	
1. Develop a research plan for each (new) question	++
2. Determine a set of core questions that are asked in each wave so that time trends can be identified	+
3. Keep (most of) the questions so that the questionnaire development and translation phase can be skipped	+
4. Use and develop validated scales in the questionnaire	+++
5. Collect basic information about the MM and ER respondent (age, educational level, ...)	+
Sampling	
6. Document yield rates/non-response by country, sector and firm size so that sampling design can be optimised	+
7. Document the share of employees employed by firms with at least 10 employees	+
Fieldwork	
8. Provide sufficient time between the pilot and mainstage fieldwork so that the lessons learnt can be incorporated	++
9. Develop a strategy to deal with the low number of ER interviews (options: set ER targets by country; switch to a mixed-mode/phone survey; no longer conduct ER interviews; conduct a separate ER and MM survey)	+++
Monitor data quality	
10. Document the threshold of the overall quality index used to discard low-quality observations	+
11. Program an additional consistency check comparing the number of employees reported in the screener with the number reported by the MM	+
Weighting	
12. Consider using the paradata to correct for non-response bias	++
Data and result dissemination	
13. Encourage increased data use	++
External validity	
14. Verify and explain time trends of key variables between the ECS 2013 and 2019	++
The quality assurance plan	
15. Ensure that the QAP guides survey processes (and not vice versa)	++

3. [Questionnaire development] Keep (most of) the questions so that the questionnaire development and translation phase can be simplified (*medium impact; easy to implement*)

A critical, but time-consuming step in the ECS is the questionnaire development and the questionnaire translation, including the advance translatability assessment and the cognitive interviews. As discussed, the ECS 2019 followed state-of-the-art translation procedures. It might therefore be considered to use a similar questionnaire in the next wave so that the questionnaire development and translation phase can be skipped or, at least, shortened. This would simplify the survey's lifecycle.

This recommendation aligns well with the previous one which recommended to select a core set of questions. Rather than revising the entire questionnaire, it might be useful to focus on a few questions that are considered less reliable or to focus on developing a few new questions to measure concepts not yet covered by the ECS.

4. [Questionnaire development] Use and develop validated scales in the questionnaire (*medium impact; difficult to implement*)

Cross-country surveys require validated scales so that questions used to measure concepts are interpreted in the same way by different respondents and across regions and cultures. The ECS intends to measure complex concepts such as teamwork or skills (mis)match. It might be considered to document (1) whether validated scales exist in the literature and (2) whether the ECS adopted a validated scale. Questions measuring the same concept in the ECS and EWCS could also be harmonised. If no validated scale exists, EF/CF could consider commissioning research to validate a scale. By listing and documenting validated scales, EF/CF could set standards that might also be adopted by national or project-specific surveys. This would facilitate data harmonisation in the longer term.

5. [Questionnaire development] Collect basic information about the MM and ER respondent (*low impact, moderately difficult to implement*)

Respondents' characteristics can influence responses and, hence, data quality. This cannot be evaluated because the current questionnaire collects almost no information about the respondent. We propose to include a few additional questions to capture respondents' characteristics, e.g. age, educational level, job tenure. Including these questions seems easy but may require dropping a few other questions. Rather than including several questions related to respondents' characteristics at once, we proposed to experiment with a few questions, test which respondents' characteristics cause measurement error and only include those questions that really matter.

6. [Sampling] Document yield rates/non-response by country, sector and firm size so that sampling design can be optimized (*medium impact, easy to implement*)

The ECS 2019 was one of the first nationally representative EU-wide push-to-web surveys and provides valuable insights in the do's and don'ts of a push-to-web survey. These insights are also informative for other surveys for which a change of mode is currently being considered. One important practical consideration are the yield rates. Yield rates are determined by the proportion of successful screener interviews and the proportion of respondents who complete the online survey once they have received an online invitation. As documented in Section 2.3, yield rates were particularly low in the ECS 2019, mainly because push-to-web surveys introduce an additional step compared to conventional phone surveys. It would be useful to clearly document yield rates by

country, sector and firm/establishment size and to make this information publicly available. This could inform the decision whether to implement a push-to-web survey as the push-to-web survey worked well in some (but certainly not all) countries. In addition, it would help to design the sampling strategy and to decide on the initial size of the sample that should be released.

7. [Sampling] Document the share of employees employed by firms with at least 10 employees (*low impact, easy to implement*)

The ECS is representative for the population of establishments employing at least 10 people in the NACE sectors B to N, R and S. The share of the working-age population employed by establishments covered by the ECS varies by country. To facilitate the interpretation of the results and to make it easier to compare the ECS findings to findings from surveys representative of employees (e.g. the EWCS) or the working-age population (e.g. LFS), it might be considered to document for each country the share of workers employed by establishments covered by the ECS.

8. [Fieldwork] Provide sufficient time between the pilot and mainstage fieldwork so that the lessons learnt can be incorporated (*medium impact; moderately difficult to implement*)

One of the strengths of the ECS 2019 was the pilot survey, which allowed Ipsos/EF/CF to test and optimise survey choices (e.g. should a screener interview be conducted with the manager or with the first person that picks up the phone; what is the expected yield rate; ...) and served as a dress rehearsal. This was all the more important given the implementation of the novel push-to-web survey methodology. Although cumbersome, conducting pilots in future waves is worth the effort given the positive effects on data quality during mainstage fieldwork. Sufficient time needs, however, to be provided between the pilot and mainstage fieldwork so that lessons learnt can be incorporated in mainstage fieldwork. Given that mainstage fieldwork lasts for about 20 weeks and has to end before the summer break (mainstage fieldwork runs from early February until June), pilot fieldwork should ideally be organised in September and October so that the three remaining months prior to the start of mainstage fieldwork can be used to incorporate the lessons learnt and prepare mainstage fieldwork. This also implies that the source questionnaire is ideally translated before the summer break. It may also be considered to specify in the contract between EF/CF and the market research company that the contract can be renegotiated after the pilot so that there is room to substantially revise mainstage fieldwork after the pilot.

9. [Fieldwork] Develop a strategy to deal with the low number of ER interviews (*high impact, difficult to implement*)

One of the limitations of the ECS 2019 is the low number of achieved ER interviews. Only 3,095 ER were interviewed compared to 6,379 in the ECS 2013. The number of establishments where both a MM and ER is interviewed is 1,835. As a result, the ECS 2019 contains too few ER observations to conduct detailed analyses related to employee representation at the country level, or to compare responses from the MM and ER working in the same establishments. These issues deserve considerable attention, but there are no easy solutions. We suggest a few options which could be further explored.

a) Set targets by country with respect to the number of ER interviews

Targets with respect to the number of ER interviews would ensure a sufficient number of interviews. However, this approach would be very expensive and does not work in smaller countries (e.g. Malta and Cyprus), where the ECS basically contacted all eligible establishments. As documented in Section 2.3, considerable efforts (email reminders,

additional CATI interviews) were made to increase the number of ER interviews. The main problem is that 68% of the managers refused to provide ER contact details during the screener interview. Reducing the refusal rate might be possible by, for instance, closely collaborating with local trade unions who might be able to provide ER contact details or could reach out to the managers. This will, however, be a cumbersome and expensive process. Moreover, it does not avoid the issue of selection bias which leads to a non-random selection of ER. We therefore do not recommend this approach.

b) Switch to a mixed-mode/phone survey

The technical and fieldwork report suggests considering a mixed-mode survey, in which either non-respondents are followed up over phone and could then complete the interview over phone or respondents can choose the interview mode (phone or web) during the screener interview. A mixed-mode survey has the advantage that the push-to-web methodology can be kept in those countries where it worked well, and be replaced by a phone survey in countries where the conversion rates at the CAWI stage were particularly low. The disadvantage is that the survey mode will no longer be the same for all respondents in all countries, which induces 'mode effects'. It is difficult to assess to which extent mode effects affect data quality. The advantage of more ER interviews has therefore to be balanced against the disadvantage of introducing mode effects. Alternatively, it could be considered to switch back to a phone survey as in the previous waves.

c) No longer conduct ER interviews in the ECS

A simple solution is to stop interviewing the ER. This approach would simplify the survey and substantially reduce the costs. This approach seems appropriate if EF/CF decide to continue with the push-to-web methodology and if no effective approach is found to reach the ER. It seems better to decide not to interview the ER from the start than to continue spending resources without much results.

d) Conduct a separate ER and MM survey

It could also be considered to implement the current ECS only for the managers, while conducting a separate survey for the ER. A feasibility study could look into the benefits and costs of this approach.

10. [Monitor data quality] Document the threshold of the overall quality index used to discard low-quality observations (low impact, easy to implement)

EF/CF developed a script to flag low-quality observations and then dropped observations below a certain quality score threshold. We recommend documenting this threshold. In addition, the quality score could be added to the final dataset so that data users can verify whether their findings are sensitive to including or excluding certain observations.

11. [Monitor data quality] Program an additional consistency check comparing the number of employees reported in the screener with the number reported by the MM (medium impact, easy to implement)

The establishment's size is a key variable which determines, amongst others, the sampling weight. Information on the establishment size is often included in the sampling frame and is reported during the screener interview as well as the MM interview. The number of employees reported during the screener can, however, deviate substantially from the number reported by the MM. An additional consistency check could be scripted in the MM interview showing a

warning message when the reported number of employees differs substantially from the number reported in the screener. As such, MM are prompted to verify their estimate. This could reduce typing error and outliers and would improve data quality.

12. [Weighting] Consider using the paradata to correct for non-response bias (*low impact, moderately difficult to implement*)

The current survey already collects detailed paradata. The survey literature suggests that paradata can be used to correct non-response bias. This could complement the current strategy which corrects non-response bias using data on the sector, establishment size and region. We recommend commissioning a feasibility study to examine whether using paradata to correct for non-response bias improves the survey representativeness.

13. [Data and results dissemination] Encouraged increased data use (*high impact, moderately difficult to implement*)

The limited number of data downloads and the limited use of the ECS data in policy reports and academic papers is in our view a second key concern that needs to be addressed before implementing a next wave. It is important to understand (i) whether the data are indeed underused by the academic community (as our analysis and a previous assessment suggest) by, among others, comparing the use of the ECS to the use of similar datasets (e.g. the ESENER-2 dataset); (ii) why the data are not frequently used and (iii) which strategies can be put in place to increase the visibility and use of the data.

14. [External validity] Verify and explain time trends of key variables between the ECS 2013 and 2019 (*medium impact, moderately difficult to implement*)

We compared the prevalence of ER according to the ECS 2019 and the ECS 2013 and attempted to explain the differences by the decreasing trade union density. While the ECS is not designed to identify time trends, the findings do, however, raise concerns about the data quality. For instance, the reported prevalence of ER was about 25 percentage points lower in Denmark and the Netherlands in 2019 compared to 2013 and decreased (quite substantially) in nearly all countries. Our analyses are only preliminary and do not explain these differences. More careful analyses are recommended to shed light on whether specific survey design issues explain this sharp decrease (e.g. the types of ER identified in the survey, the switch to the push-to-web methodology). These lessons can then inform the next wave. Similar analyses could also be conducted for other key variables.

15. [The Quality Assurance Plan] Ensure that the QAP guides survey processes (and not vice versa) (*medium impact, moderately difficult to implement*)

Continuous quality monitoring throughout the survey's lifecycle is fundamental. The Quality Assurance Plan (QAP) is one element of quality monitoring. Such a plan attempts to strike the right balance between the need for specific, measurable objectives without overburdening the market research company and EF/CF with reporting on these objectives. It is important that the QAP guides survey processes and is not considered an administrative obligation that has to be completed once a month. While we do not feel that this was the case in this wave (due to close monitoring of the plan throughout the survey's implementation and based on the fieldwork contractor's own assessment), it remains a risk due to the high number of indicators. The intense discussion between EF/CF and the fieldwork contractor in an early phase of the survey and the designated Delivery Manager at Ipsos in

charge of updating the QA indicators helped to closely monitor the QAP. These practices could also be adopted in the next waves and in other surveys.

In addition it might be considered to, if possible, reduce the number of indicators and simplify the QAP. To do so, one could consider reducing the number of quality dimensions (e.g. only one QA indicator relates to 'Relevance and Timeliness', so this quality dimension could be dropped). Finally, it is worth emphasising that reaching all QA indicators does not necessarily result in high-quality data since (1) not all issues that could go wrong are captured by the QA indicators and (2) it is not always possible to tell whether targets are feasible before the start of the survey (e.g. targets related to the number of ER interviews).

References

All Eurofound publications are available at www.eurofound.europa.eu

- Biemer, P. P., Chen, P., and Wang, K. (2013), 'Using level-of-effort paradata in non-response adjustments with application to field surveys', *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 176(1), 147-168.
- Conway, K., Acquadro, C., and Patrick, D. L. (2014), 'Usefulness of translatability assessment: results from a retrospective study', *Quality of Life Research*, 23(4), 1199-1210.
- Couper, M. P., Tourangeau, R., and Kenyon, K. (2004), 'Picture this! Exploring visual effects in web surveys', *Public Opinion Quarterly*, 68(2), 255-266.
- de Bruin, W. B., Van der Klaauw, W., Topa, G., Downs, J. S., Fischhoff, B., and Armantier, O. (2012), 'The effect of question wording on consumers' reported inflation expectations', *Journal of Economic Psychology*, 33(4), 749-757.
- ESS (2019), *Quality Assurance Framework of the European Statistical System. Version 2.0*. Retrieved from <https://ec.europa.eu/eurostat/documents/64157/4392716/ESS-QAF-V1-2final.pdf/bbf5970c-1adf-46c8-afc3-58ce177a0646>
- Eurofound (2011), *Quality Assessment of the 2nd European Company Survey*, Dublin.
- Eurofound (2013), *Quality Assessment of the third European Company Survey*, Dublin.
- Eurofound (2015), *Third European Company Survey – Overview report: Workplace practices – Patterns, performance and well-being*, Publications Office of the European Union, Luxembourg.
- Eurofound (2017a), *Feasibility study regarding methodology, design and mode of the European Company Survey. Task 1: Methodological lessons learnt from the last European Company Survey (3rd ECS, 2013)*, unpublished.
- Eurofound (2017b), *Feasibility study regarding methodology, design and mode of the European Company Survey. Task 2: sampling modes and frames*, unpublished.
- Eurofound (2018), *European Quality of Life Survey 2016: Quality Assessment*, Eurofound working paper, Dublin.
- Eurofound (2019), *Skills utilisation: Definition, theories, approaches and measures*, Eurofound working paper, Dublin
(<https://www.eurofound.europa.eu/sites/default/files/wpef18048.pdf>)
- Eurofound and Cedefop (2018a), *European Company Survey 2019. Pilot report*, unpublished.
- Eurofound and Cedefop (2018b), *European Company Survey 2019. Cognitive Interview Report*, unpublished.
- Eurofound and Cedefop (2020a), *European Company Survey 2019: Technical and fieldwork report*, European Company Survey 2019 series, Eurofound working paper, Dublin
(<https://www.eurofound.europa.eu/publications/report/2020/european-company-survey-2019-workplace-practices-unlocking-employee-potential#wp-101978>).
- Eurofound and Cedefop (2020b), *European Company Survey 2019: Quality control report*, European Company Survey 2019 series, Eurofound working paper, Dublin
(<https://www.eurofound.europa.eu/publications/report/2020/european-company-survey-2019-workplace-practices-unlocking-employee-potential#wp-101982>).
- Eurofound and Cedefop (2020c), *European Company Survey 2019: Translation report*, European Company Survey 2019 series, Eurofound working paper, Dublin

Disclaimer: This working paper has not been subject to the full Cedefop/Eurofound evaluation, editorial and publication process.

- (<https://www.eurofound.europa.eu/publications/report/2020/european-company-survey-2019-workplace-practices-unlocking-employee-potential#wp-101980>).
- Eurofound and Cedefop (2020d), *European Company Survey 2019: Sampling and weighting report*, European Company Survey 2019 series, Eurofound working paper, Dublin (<https://www.eurofound.europa.eu/publications/report/2020/european-company-survey-2019-workplace-practices-unlocking-employee-potential#wp-101981>).
- Eurofound and Cedefop (2020e), *European Company Survey 2019: Data editing report*, European Company Survey 2019 series, Eurofound working paper, Dublin (<https://www.eurofound.europa.eu/publications/report/2020/european-company-survey-2019-workplace-practices-unlocking-employee-potential#wp-101979>).
- Eurofound and Cedefop (2020f), *European Company Survey 2019: Coding report*, European Company Survey 2019 series, Eurofound working paper, Dublin (<https://www.eurofound.europa.eu/publications/report/2020/european-company-survey-2019-workplace-practices-unlocking-employee-potential#wp-101983>).
- Gaziano, C. (2005), 'Comparative analysis of within-household respondent selection techniques', *Public Opinion Quarterly*, 69(1), 124-157.
- Groves, R. M. (2006), 'Nonresponse rates and nonresponse bias in household surveys', *Public Opinion Quarterly*, 70(5), 646-675.
- Groves, R. M., Fowler Jr, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., and Tourangeau, R. (2011), *Survey methodology*, John Wiley and Sons, Hoboken.
- Groves, R. M., Lyberg, L. (2010), 'Total survey error: Past, present, and future', *Public Opinion Quarterly*, 74(5), 849-879.
- Harkness, J. A. (2007), 'Improving equivalence in translations', in Jowell, R., Roberts, C., Fitzgerald, R., and Gillian, E. (eds) *Measuring attitudes cross-nationally: Lessons from the European Social Survey*, Sage, London.
- Harkness, J. A., Braun, M., Edwards, B., Johnson, T. P., Lyberg, L. E., Mohler, P. P., and Smith, T. W. (2010), *Survey methods in multinational, multiregional, and multicultural contexts* (Vol. 552), John Wiley & Sons.
- Lyberg, L. (2012), 'Survey quality', *Survey Methodology*, 38(2), 107-130.
- National Research Council (2013), *Nonresponse in social science surveys: A research agenda*, National Academies Press, Washington DC.
- Rossi, P. H., Wright, J. D., and Anderson, A. B. (2013), *Handbook of survey research*, Academic Press, London.
- Saris, W. E., & Gallhofer, I. N. (2014), *Design, evaluation, and analysis of questionnaires for survey research*, John Wiley and Sons, Hoboken.
- Scherpenzeel, A. C., Saris, W. E. (1997), 'The validity and reliability of survey questions: A meta-analysis of MTMM studies', *Sociological Methods & Research*, 25(3), 341-383.
- Tourangeau, R., Conrad, F. G., & Couper, M. P. (2013), *The science of web surveys*, Oxford University Press, Oxford.
- Van de Walle, S., Van Ryzin, G. GF. (2011), 'The order of questions in a survey on citizen satisfaction with public services: lessons from a split-ballot experiment', *Public Administration*, 89(4), 1436-1450.
- Vander Elst, T., De Witte, H., and De Cuyper, N. (2014), 'The Job Insecurity Scale: A psychometric evaluation across five European countries', *European Journal of Work and Organizational Psychology*, 23(3), 364-380.

Appendix 1

Table 19: Sampling frames used in 2013 and 2019

	Sampling frame ECS 2013	Sampling frame ECS 2019	Same sampling frame in 2013 and 2019
<i>Austria</i>	Dun & Bradstreet	HEROLD Marketing CD / MDOOffline professional	Yes
<i>Belgium</i>	Infobel	Graydon	No
<i>Bulgaria</i>	Bulgarian National Statistical institute	Dun & Bradstreet (centrally sourced)	No
<i>Croatia</i>	Annual Financial Statements Registry (RGFI), Financial Agency (FINA)	Annual Financial Statements Registry (RGFI), Financial Agency (FINA)	Yes
<i>Cyprus</i>	Business Register (ΜΗΤΡΩΟ ΕΠΙΧΕΙΡΗΣΕΩΝ)	Business Register (ΜΗΤΡΩΟ ΕΠΙΧΕΙΡΗΣΕΩΝ)	Yes
<i>Czechia</i>	Bisnode Albertina	Bisnode Albertina	Yes
<i>Denmark</i>	Solidet	Bisnode	No
<i>Estonia</i>	Kreddiinfo	e-Business Register	No
<i>Finland</i>	Fonecta	Bisnode	No
<i>France</i>	LBM Direct	Direct Fichier	No
<i>Germany</i>	Dun & Bradstreet	Heins & Partner	No
<i>Greece</i>	ICAP directory	ICAP directory	Yes
<i>Hungary</i>	KSH	Bisnode	No
<i>Ireland</i>	Bill Moss	Bill Moss	Yes
<i>Italy</i>	Dun & Bradstreet	Cribis	No
<i>Latvia</i>	Business register of Central Statistical Bureau	Lursoft	No
<i>Lithuania</i>	Creditinfo	Creditinfo	Yes
<i>Luxembourg</i>	Editus	Editus	Yes
<i>Malta</i>	Employment & Training Corporation	Dun & Bradstreet (centrally sourced)	No
<i>Netherlands</i>	Handelsregister	Handelsregister	Yes
<i>Poland</i>	PCM	Dun & Bradstreet (centrally sourced)	No
<i>Portugal</i>	Informa D&B	Informa D&B	Yes

Disclaimer: This working paper has not been subject to the full Cedefop/Eurofound evaluation, editorial and publication process.

	Sampling frame ECS 2013	Sampling frame ECS 2019	Same sampling frame in 2013 and 2019
<i>Romania</i>	Lista Firmelor din Romania	Lista Firmelor din Romania	Yes
<i>Slovakia</i>	Bisnode Albertina	Bisnode Albertina	Yes
<i>Slovenia</i>	Ipis	Bisnode Slovenija	No
<i>Spain</i>	Shober	Informa D&B	No
<i>Sweden</i>	Parad	Bisnode Sverige (PARAD)	No
<i>United Kingdom</i>	Dun & Bradstreet	Experian	No

Source: (Eurofound and Cedefop, 2020a)

Table 20: Target and achieved sample size

	Reference N (prior to pilot study)	Final N (targets for mainstage fieldwork)	Achieved sample size (after dropping observations of low- quality)	% of targets
Austria	1,000	1,000	1,010	101%
Belgium	1,000	1,000	1,011	101%
Bulgaria	1,000	1,000	1,024	102%
Croatia	500	500	560	112%
Cyprus	250	250	122	49%
Czechia	1,000	1,000	904	90%
Denmark	1,000	1,000	1,011	101%
Estonia	500	500	501	100%
Finland*	1,000	1,000	1,032	103%
France	1,500	1,500	1,360	91%
Germany	1,500	1,000	711	71%
Greece	500	500	501	100%
Hungary	1,000	1,000	1,087	109%
Ireland	500	250	300	120%
Italy	1,500	1,500	1,498	100%
Latvia	500	500	514	103%
Lithuania	500	500	510	102%
Luxembourg	250	250	237	95%
Malta	250	250	145	58%
Netherlands	1,000	1,000	1,030	103%
Poland	1,500	1,150	842	73%
Portugal	1,000	1,000	973	97%
Romania	1,000	1,000	815	82%
Slovakia	500	350	361	103%
Slovenia	500	500	556	111%
Spain	1,500	1,500	1,477	98%
Sweden	1,000	1,000	1,080	108%
United Kingdom*	1,500	700	697	100%
Total EU28	24,750	22,700	21,869	96%

Source: (Eurofound and Cedefop, 2020a)

Note: Coloured boxes indicate countries where the sample size was reduced after the pilot because the yield rate in the pilot was lower than expected. * In Finland and the UK the agreed targets after the pilot were initially smaller, but were increased again because the yield rates during mainstage fieldwork were better than expected

Table 21: Reasons for unsuccessful screener interviews (EU average) by country

	All Countries	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Share of successful screeners	14%	12%	21%	22%	27%	12%	5%	23%	22%	32%	12%	14%	25%	13%	14%	14%	35%	48%	17%
Screener interview not successful	379762	16949	11295	11886	4395	3272	34348	9030	3524	4954	20664	25223	7293	18085	7233	21671	3005	1750	4808
Remains in current sample	7%	7%	7%	19%	23%	0%	9%	5%	1%	0%	5%	19%	13%	0%	4%	4%	0%	7%	24%
Breakoff during screener	1%	1%	3%	2%	4%	2%	1%	1%	2%	1%	0%	2%	1%	1%	1%	2%	4%	5%	1%
Appointment	5%	0%	2%	0%	12%	7%	0%	10%	0%	2%	4%	1%	18%	12%	27%	5%	0%	1%	1%
Establishment not eligible or no information on key variables (type and size of establishment)	10%	22%	8%	13%	6%	4%	2%	8%	3%	9%	10%	4%	20%	3%	3%	12%	24%	9%	22%
Call failed – recall	24%	6%	20%	18%	17%	23%	19%	20%	41%	29%	37%	9%	4%	25%	20%	32%	18%	13%	10%
Call failed – unusable	11%	6%	8%	29%	14%	7%	8%	13%	29%	9%	18%	12%	7%	8%	9%	14%	15%	7%	23%
Refusal	42%	57%	51%	17%	25%	52%	63%	43%	25%	50%	25%	52%	37%	51%	36%	32%	39%	56%	19%
Language barrier	0%	0%	2%	1%	0%	4%	0%	1%	0%	1%	0%	0%	0%	1%	0%	0%	0%	2%	1%

	All Countries	19	20	21	22	23	24	25	26	27	28
Share of successful screeners	14%	19%	18%	4%	14%	14%	14%	45%	11%	20%	18%
Screener interview not successful	379762	1462	11467	53626	15725	17830	9193	2042	32468	10047	16517
Remains in current sample	7%	0%	5%	2%	19%	8%	26%	2%	4%	0%	0%
Breakoff during screener	1%	0%	2%	1%	1%	2%	1%	5%	1%	3%	1%
Appointment	5%	0%	3%	11%	0%	4%	0%	1%	3%	2%	0%
Establishment not eligible or no information on key variables (type and size of establishment)	10%	94%	14%	18%	8%	4%	2%	8%	8%	7%	7%
Call failed – recall	24%	0%	10%	34%	1%	22%	27%	16%	34%	27%	45%
Call failed – unusable	11%	1%	16%	7%	5%	9%	11%	14%	8%	15%	13%
Refusal	42%	5%	50%	27%	65%	49%	34%	52%	41%	45%	34%
Language barrier	0%	0%	1%	0%	1%	1%	0%	1%	0%	1%	0%

Source: Appendix 2

Note: Numbers correspond to EU28 Member States in alphabetical order (1: Austria; 2: Belgium; 3: Bulgaria; 4: Croatia; 5: Cyprus; 6: Czechia; 7: Denmark; 8: Estonia; 9: Finland; 10: France; 11: Germany; 12: Greece; 13: Hungary; 14: Ireland; 15: Italy; 16: Latvia; 17: Lithuania; 18: Luxembourg; 19: Malta; 20: Netherlands; 21: Poland; 22: Portugal; 23: Romania; 24: Slovakia; 25: Slovenia; 26: Spain; 27: Sweden; 28: UK)

Disclaimer: This working paper has not been subject to the full Cedefop/Eurofound evaluation, editorial and publication process.

Figure 19: Correlation between OECD's indicator 'trade union density' and the ECS indicator 'presence of an ER'

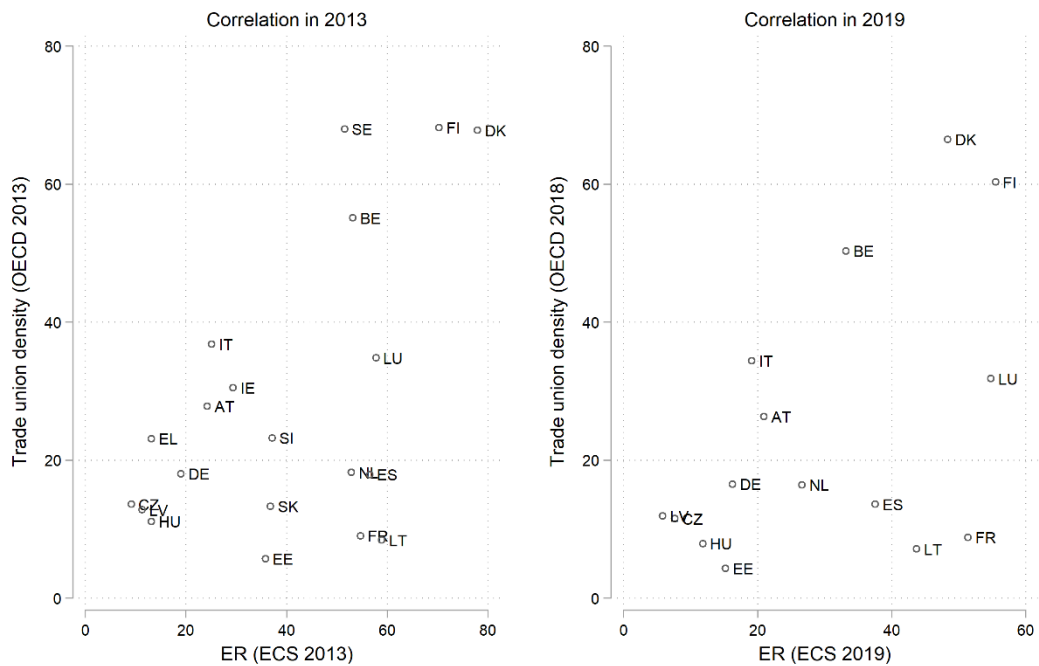


Table 22: Selection bias: ER contact details shared by the MM in the screener interview (logistic regression)

	Outcome variable: ER contact details shared	
General attitude of the ER (base level: very constructive)		
Fairly constructive	-0.313***	-0.324***
	(0.0855)	(0.0918)
Not very constructive	-0.537***	-0.481***
	(0.119)	(0.129)
Not at all constructive	-0.611**	-0.619**
	(0.260)	(0.272)
Establishment size (log)	0.0215	0.0260
	(0.0301)	(0.0333)
Controls for countries	No	Yes
Observations	3,022	3,020

Source: author calculations on ECS 2019 management dataset

Note: Variable general attitude (var: 'eratt'), Standard errors in parentheses. ***, **, * denote statistical significance at the 1%, 5% and 10% level, respectively

Table 23: Using the length of the screener interview (minutes) to predict the probability of completing the MM questionnaire (after a successful screener)

	Outcome: MM survey completed (after a successful screener)
Length of the screener interview (minutes)	0.0403***
	(0.00634)
(Length of the screener interview)²	-0.00145***
	(0.000255)
Establishment size (base level: no size stratum)	
Small	0.427***
	(0.0902)
Medium	0.404***
	(0.0915)
Large	0.264***
	(0.0937)
Sector (baselevel: production)	
Construction	-0.108***
	(0.0314)
Services	-0.0732***
	(0.0201)
Controls for countries	Yes
Observations	65,052

Source: author calculations on ECS 2019 management dataset

Note: Standard errors in parentheses. ***, **, * denote statistical significance at the 1%, 5% and 10% level, respectively

Appendix 2

Table 24: Unpublished reports and other material used in figures and tables that can be made available upon request

Tables and Figures	Unpublished reports	Information also available
Figure 1	Internal documents and Eurofound and Cedefop (2020a)	
Figure 2	Presentation from Eurofound and Cedefop during the kick-off meeting June 25, 2019, Leuven	European Company Survey 2019 overview report: workplace practices unlocking employee potential
Table 3	Excel file '161016 - ECS 2019 - Questionnaires MM and ER - Final version MAINSTAGE - 1 March 2019'	Questionnaires available: https://www.eurofound.europa.eu/surveys/2019/european-company-survey-2019
Table 4	Excel file '161016 - ECS 2019 - Questionnaires MM and ER - Final version MAINSTAGE - 1 March 2019'	Questionnaires available: https://www.eurofound.europa.eu/surveys/2019/european-company-survey-2019
Figure 3	Management questionnaire	https://www.eurofound.europa.eu/sites/default/files/ef_survey/field_ef_documents/ecs_2019_mm_questionnaire_for_publication_clean_14_july_2020.pdf
Figure 6	Technical Specification. Preparation and Implementation of the 4th European Company Survey (ECS), Table 8	
Figure 7	Technical Specification. Preparation and Implementation of the 4th European Company Survey (ECS), Table 3.2 and Table 8	
Table 8	RAG report week 20	
Table 21	RAG Reports Week 2 – 20	

WPEF19017

The European Centre for the Development of Vocational Training (Cedefop) is the European Union’s reference centre for vocational education and training. It provides information on and analyses of vocational education and training systems, policies, research and practice. Cedefop was established in 1975 by Council Regulation (EEC) No 337/75.

The European Foundation for the Improvement of Living and Working Conditions (Eurofound) is a tripartite European Union Agency established in 1975. Its role is to provide knowledge in the area of social, employment and work-related policies according to Regulation (EU) 2019/127.