InGRID

Integrating expertise in inclusive growth

www.inclusivegrowth.be

Milestone 21.13b

EVALUATION OF THE EWCS 2010 AND THE MEASUREMENT OF JOB QUALITY WITH THE TOTAL SURVEY ERROR APPROACH

Anina Vercruyssen & Guy Van Gyes

February 2017

This project has received funding from the European Union's Seventh Programme for Research, Technological Development and Demonstration under Grant Agreement No 312691

This report constitutes Milestone 21.13b 'Methods of identify vulnerable groups', for Work Package 21 of the InGRID project.

February 2017

General contact: inclusive.growth@kuleuven.be
p.a. InGRID
HIVA - Research Institute for Work and Society
Parkstraat 47 box 5300, 3000 LEUVEN, Belgium

For more information guy.vangyes@kuleuven.be

Please refer to this publication as follows:
Vercruyssen, A., & Van Gyes, G. (2017). *Evaluation of the EWCS 2010 and the measurement of job quality with the total survey error approach* (working paper M21.13b). Leuven: InGRID.

This publication is also available via http://www. Inclusivegrowth.be/project-output

This publication is part of the InGRID project, this project has received funding from the European Union's Seventh Programme for Research, Technical Development and Demonstration under Grant Agreement No 312691.

Integrating expertise in inclusive growth

European policy-oriented research can and must deliver useful contributions to tackle the Europe 2020 challenges of Inclusive Growth. Key tools in this social sciences research are all types of data earning statistics, administrative social data, labour market data, surveys on quality of live or working conditions, policy indicators. The project aims to integrate and optimise these existing European data infrastructures and accompanying expertise.

Contents

Introduction

The general objective of the InGRID project is to integrate and innovate existing but distributed European social sciences research infrastructures on 'poverty and living conditions' and on 'working conditions and vulnerability'. InGRID as research infrastructure project serves the social sciences community that aspires to make an evidence-based contribution to the European policy challenge of inclusive growth. It is a social sciences community that focuses on social in/exclusion, vulnerability-at-work and related social and labour market policies from a European comparative perspective. It is an interdisciplinary field of poverty research, labour studies, policy analysis and social statistics. Key tools in this social science research are all types of data: statistics on earnings, administrative social data, labour market data, surveys on quality of life or working conditions, and policy indicators.

In the field of comparative research on working conditions in Europe, the European Working conditions Survey (EWCS) is a key resource.

Since its launch in 1990 the European Working Conditions Survey has provided an overview of working conditions in Europe in order to:

- assess and quantify working conditions of both employees and the self-employed across Europe on a harmonised basis;
- analyse relationships between different aspects of working conditions;
- identify groups at risk and issues of concern as well as of progress;
- monitor trends by providing homogeneous indicators on these issues.

As such, it contributes as main comparative data source to European policy development in particular on quality of work and employment issues. An evaluation study of the fifth 2010 EWCS showed that from 2010 through to the end of 2013 the survey has been used in over 150 European policy reports or working documents. An additional review identified some 50 or more academic articles published during the same period (Eurofound, 2014). The value from the perspective of the InGRID research and policy community can in others worth not be underestimated.

The survey is organised and provided by the European Foundation for the Improvement of Living and Working Conditions. This European agency - short name Eurofound - is based in Dublin and is one of the oldest agencies set up by the European Union. It exist for more than 40 years and is governed by a tripartite structure, involving governments and social partners (employers' organisations and trade unions).

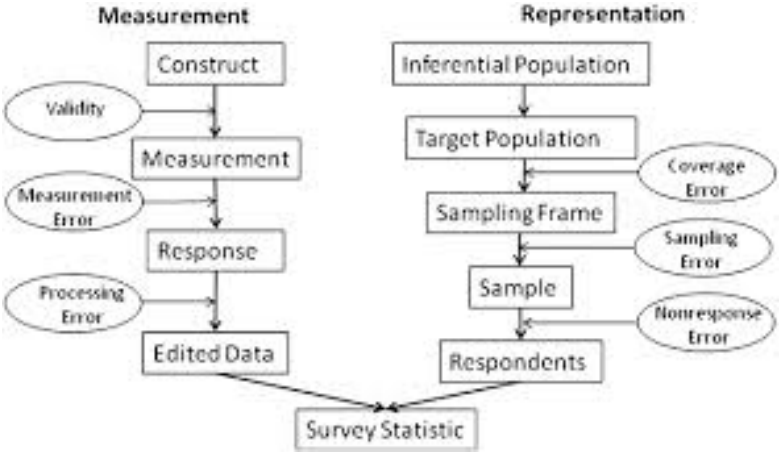
Due to the high importance for the policy-oriented research community that InGRID as infrastructure wants to serve, the InGRID partner HIVA-KU Leuven works closely together with the EWCS organiser and data provider to provide transnational access to the data and to guide, coach and train people to use this data in the best/better way.

Taking this highly valued central position of the EWCS data in the InGRID community as a starting point, the current paper implies as an exercise - positive, but critical - the total survey error evaluation framework to the EWCS data collection. The question whether there are (maybe) methodological elements of improvement to detect, is tackled from this methodological ideal type quality framework. It is about showing a 'loved' one, where its beauties lies or not, by holding a 'strong' mirror up.

The total survey error (TSE) approach is a strong guideline for evaluating the quality of survey data and measurement instruments. Key works on TSE distinguish two major components: representation

error and measurement error (Bautista, 2012; Groves *et al.*, 2004; Groves & Lyberg, 2010, see Figure 1.1). The former refers to how well the target population or target groups are represented in the survey data, the latter concentrates on how well the research questions obtain accurate answers. We evaluate the EWCS data from 2010 with this TSE-approach and will explain its subcomponents alongside of the application of the framework on the survey data.

Figure 1.1 TSE components proposed by Groves & Lyberg (2010)



1. Evaluation of the representativity in EWCS 2010

1.1 Inferential population

The inferential population is the population about which the survey wants to be able to draw conclusions and discuss implications. In the case of EWCS, the survey wants to investigate the working conditions of employees and the self-employed across Europe. In the 2010 edition of the survey, face-to-face interviews were organised by Gallup Europe in the EU-27, Norway, Croatia, the former Yugoslav Republic of Macedonia, Turkey, Albania, Montenegro and Kosovo (EWCS, 2013). EWCS 2010 did not only include all Member States of the European Union in its inferential population, the delimitation of the geographical area also already included (potential) future member states. Each country has a national fieldwork partner of Gallup Europe (list available in the technical report, see Gallup Europe, 2013) to organise the interviews in each (potential future) member state. The inferential population is clearly delimited.

1.2 Target population

The inferential population usually gets reduced to a target population, mostly because of practical considerations. The target population of EWCS 2010 has been defined as people living in the 34 aforementioned countries, aged 15 or older (16 or older in Spain, the UK and Norway) that were in employment at the time of the survey (minimum 1 hour of work for pay or profit). People (temporarily) living in institutions, such as prison or medical institutions, tend to be excluded from the population of interest for surveys because they are not living in regular households and cannot easily be interviewed. This was also the case for EWCS 2010 (see Gallup Europe, 2013a, p. 12-13). However, prisoners can also have small jobs within the institution and temporarily hospitalised people can still belong to the working population. This could form a first, though extremely small, source of systematic error or bias.

For the ease of conducting a face-to-face survey such as EWCS, it can also be opted to exclude remote parts of a country or region to limit the travel time and expenses of the interviewers sent out to collect the data. This forms another reduction of the inferential population to a target population. For EWCS 2010, some overseas parts and remote islands, such as the Canaries belonging to Spain, were excluded (Gallup Europe, 2013d). As such, these islanders are not represented in the target population nor in the survey. Sweden and Norway, however, resorted to first contact by telephone to reduce the drain on their resources to contact people at more remote locations. Extrapolating from this particular case, EWCS seems to exclude (quicker) European people with a paid job when they live in more remote locations. Given that these areas are chronically under-surveyed, it might actually be even more necessary to investigate the working conditions and implementation of (job-related) policies in such remote areas as they escape scrutiny too often. Although using mixed survey modes come with its own specific set of implications for representativity, using CATI (telephone interview) or self-administered surveys instead of CAPI would allow eligible sample units to be included in the target population.

1.3 Sample frame and potential coverage errors

Once the target population has been delimited, an instrument is needed to identify all persons in this target population: a sample frame. For EWCS 2010, this sample frame preferably needed to be 'an updated, good quality sampling frame (register) with addresses/persons' (Gallup Europe, 2013a, p. 11). EWCS 2010 considers list-based samples of individuals as the best approach, followed by lists-based sampled of addresses or households (Gallup Europe, 2013b). An official population register with individual or household level information is the most optimal sample frame as it should only have minimal coverage error. A typical coverage error to arise in official registry data is due to the timeliness of the updates of this register: the registration of demographical changes, such as deaths and address changes, can take a while and will lead to a small proportion of misinformation. Deaths, for example, will lead to over-coverage as these persons are still part of the lists but are no longer available for an interview. An example of potential under-coverage in official registry lists are persons who have immigrated to the country but whose registration documents are still being processed.

When such registers are not available, a random route method was implemented for EWCS. A random route or random walk procedure implies that before the interviewers are sent into the field, enumerators are on the road to create a sample of addresses by starting at (randomly chosen) locations and take a walk in the area. Such a walk is sometimes regulated by strict instructions, e.g. take every second street to the right on your path. The intention of this process with strict instructions is to create equal selection probabilities, but unfortunately, this procedure does not always lead to uniformly distributed probabilities for samples of households (Bauer, 2014). For the sample of the 5th edition of EWCS '*theoretically all members of the statistical population had a known non-zero probability*' (Gallup Europe, 2013c, p. 11). Sampling sources per country can be found in the technical reports (e.g. Gallup Europe, 2013a, p. 13), showing the majority of the countries (20 out of 34) relied on random walk sampling.

Why the need to know the inclusion probabilities so precisely? This is because we need to know that everybody has a similar or calculable chance to be selected for the interview in order to let the stochastic processes enable inference of survey data to the bigger population. In other words, if some people are (systematically) missing from the sample frames, we cannot achieve perfect representation of the inferential population because certain types of people's opinions will never be heard in the survey. Under-coverage may arise e.g. for small streets that are easily overlooked or new streets that are not included on the maps yet. Mistakes or manipulation by enumerators could also lead to biased address lists; deviations from the systematic walk can lead to over-coverage by including streets that should not have been included according to the walking instructions. In cases in which the enumerator had doubts about the eligibility of the household on his random walking route, the address was listed anyway and the interviewers were sent out to attempt gaining contact. 10% of the enumerated addresses were checked (Gallup Europe, 2013d). When non-compliance to the random walk rules was discovered, the primary (geographic) sampling units were replaced. The number of times this occurred within the 10%-subsample was not specified, making it impossible to estimate the size of coverage error.

As a summary of potential coverage error in EWCS 2010, we can say that we see potential over-coverage and under-coverage in all three types of sample frames. Both registers on individual and household levels can be prone to administration issues that can lead to people being wrongfully included or excluded in the sample frame, such as deceased persons and people who have just moved. For the enumeration, mistakes can happen and simulations show that random walk methods often fail to create equal selection probabilities (Bauer, 2014), leaving the door wide open for over-coverage and under-coverage of streets and addresses. All types of sampling will also have significant portions of over-coverage with regard to the employment status of people belonging to the secondary sampling units (the households). With only the information from the address-based sampling frames, it is hard to pre-select only those with households containing individuals with a paid job. There will be quite a few households only harbouring individuals without a paid job, leading to plenty of listings of

people who do not belong to the target group of the survey. According to the technical report (Gallup Europe, 2013c), 36,402 people/households over all participating countries were identified as covering only non-workers, implying that 19% of the total sample size is actually over-coverage. Additionally, 25,583 households were classified as 'unknown eligibility', which means that for 13.3% of the total sample of households it is not known if they had at least one individual belonging to the target population that could have participated.

1.4 Sample and potential sampling error

Questioning a whole population would take a tremendous amount of resources - it would actually be almost impossible to achieve such a thing (Bethlehem, 2009). Therefore, a representative selection is taken from each sample frame to 'speak for' the inferential population. When this selected group is interviewed and when their selection probabilities are known, which is the case for EWCS according to the technical report (Gallup Europe, 2013), p. 13), their results can be inferred to make statements about the whole inferential population. Mind, the inferential population is not necessarily the total population of a nation state, as can be seen in the description of the inferential population for EWCS 2010 as well.

For EWCS, the selection of 'representative' people was achieved by a multistage, stratified, random sample when population registers were available (Gallup Europe, 2013; EWCS, 2013). Multistage and stratified implies that first bigger geographical areas were randomly chosen. In the case of EWCS, these were NUTS-regions or local statistical regions, also taking into account the level of urbanisation. In the second stage, households are drawn at random within these regions. Given that EWCS 2010 worked with households as secondary sample units, official registries that do not go beyond the household level and do not contain individual level data were equally useful. When interviewers found multiple eligible households at an address, the selection of the household was based on alphabetically ordering the surnames (of the heads of the household). When a household consists of multiple eligible persons, the instruction was to interview the individual who was the next one in the household to have a birthday (EWCS, 2013). For the latter method to work as a form of random sampling, we actually need to assume that the distribution of birthdays throughout the year equal in every participating country. However, the fieldwork periods are not spread over the whole year, meaning that mainly spring or summer-born individuals were surveyed. This is only problematic for the representativity of EWCS if there is a correlation between birthdays or birthday seasons and the type of job (quality) and/or survey participation. However, to the best of our knowledge no studies have been conducted on this matter.

EWCS 2010 sets a target number of 1,000 completed interviews per country. Some countries, such as France, were allowed to expand the 1,000 interview target to better enable national comparisons (see Gallup Europe, 2013b). To achieve those 1,000 interviews per country, the total sample size per country was, on average, set at 6.4 times as many useable addresses, meaning that on average about 6400 addresses needed to be collected or ordered per country in order to achieve 1,000 interviews. All countries managed to achieve their pre-set target, except for Spain (planned 1,500, realised 1,008). The 34 countries collected or ordered a combined number of 278,456 addresses, which can be considered as the total sample size, even though it is not clear whether all addresses were actually used.

This big ratio of useable addresses (gross sample) versus target interviews (realised sample) makes sense when taking into account all the coverage error that is generated by the sampling frames, such as the large proportion of sampled households with only non-working individuals. Additionally, not every selected sampled person will *want* to cooperate, leading to a large proportion of non-response that will prove challenging for reaching the targets if the samples are too small. Hence, it looks like a very large margin was necessary to obtain 1,000 interviews per country. Implications of non-response error are discussed in Section 1.5.

When working with samples, sampling errors can occur as well. Such errors represent the differences between the population as a whole and the sample. Drawing at random does not guarantee a perfect proportional representation of people in the sample in comparison to the whole population. Corrective weights were applied based on e.g. census data in EWCS 2010. However, given that the *de facto* population was targeted rather than the *de jure* population (see Gallup Europe 2013b, p. 10) and given that census data is often based on the *de jure* population, the choice to weigh the sample from the de factor population with information from the *de jure* population will lead to small differences, as e.g. inclusion of asylum seekers in the calculations.

Another potential source of sampling error is exchanges between primary sampling units, being between selected geographical areas. Because some of these geographical areas were less successful in terms of being able to reach the desired number of interviews, the interviewers in some regions were instructed to focus their efforts on geographical areas that had already proven to be more successful for obtaining interviews. Whereas this is a good strategy to get closer to the interview quota, it is, however, rather risky with regard to the representativity of the realised sample in comparison to the whole population. Areas where it is easier to obtain interviews may be so for reasons related to the topic of the survey, e.g. response rates may be lower in areas with lower socio-economic status (SES), areas that may also look less inviting to interviewers. If this leads to withdrawing focus from the lower SES-areas in favour of areas with a better SES, the realised sample may under-represent people with lower SES-jobs. This, in turn, would also reflect on the estimates for job quality in the survey as well, a topic for which correct estimates are important as they can be used to inform policy and policy makers. More attention should be paid to this in the next rounds of the survey.

1.5 Respondents and potential non-response error

The next level of error, which is strongly related to the sample and sometimes intertwined with sampling error, is non-response error. Of all the reports on the fifth EWCS survey, only the technical report (Gallup Europe, 2013c) contains the response and non-response rates calculated according to the AAPOR (2009) standards for this survey. The average response rate over all countries is 44.2% and ranges from 31.3% to 73.5%. While measuring a survey's success and quality using response rates is common practice in social sciences, it needs to be stressed that response rates are not the best indicators of representativity as the correlation between non-response rates and non-response bias is actually not that strong (see e.g. Groves, 2006). This means that a survey with a low response rate can theoretically be more representative than a survey with a medium/high response rate. What matters is whether non-response is systematic or random.

Non-response does not pose a big problem for the representativity of the realised sample if it is completely random. *Missing completely at random*-situations will normally not harm the survey in terms of representativity of the sampled individuals in comparison to the inferential population (see e.g. Bethlehem, 2009; Groves & Couper, 1998). When non-response is not (completely) random, however, the representativity comes under pressure. It is especially problematic when there is a correlation between the non-response and key survey variables - a situation called *Not Missing At Random* (NMAR). In the case of EWCS, systematically missing persons with e.g. a lower SES-job or with low job quality would be a NMAR-situation. It would imply a misrepresentation of this lower SES-group as well as biased estimates for scores on job quality. When certain geographical areas (the primary sample units) were abandoned in EWCS 2010, as described in Section 1.4., it is plausible that lower SES-areas and individuals with lower SES were systematically missed in the survey. This is also where sampling errors and non-response errors get intertwined in this survey.

To be able to investigate whether a non-response bias exists, we would need information on the job and job quality of the non-participants of EWCS. With the two-stage sampling, we would need information on area level and on individual/household level. However, information on non-respondents is by definition missing in surveys, as is also the case with the EWCS data. Therefore, we are

unable to determine whether and to what degree non-response bias occurs in the EWCS survey of 2010. What is known for the response processes in EWCS 2010, is that three recalls were required after the initial visit before the address could be considered as a non-contact (Gallup Europe 2013c, p11). Some countries also had records of telephone numbers they could use, but there always needed to be a face-to-face contact attempt as well. A small follow-up survey, such as a basic questionnaire approach (see e.g. Bethlehem & Kerstens, 1985), can help to gather data on non-respondents to facilitate investigating whether the survey suffers from non-response bias. A basic questionnaire is a short version of the survey, containing a minimum of relevant socio-demographic and on-topic questions. A pre-emptive basic questionnaire *during* the fieldwork, such as the PEDASKI-approach (Lynn, 2002), is even more recommendable as it does not require much extra resources and efforts because they interviewers have their back-up mini-survey during the original fieldwork whereas follow-up surveys imply extra fieldwork time and effort.

Although implementing these methods actually implies introducing a different mode of the survey with potential mode effects, it is still better to have some information on the non-respondents than none. Alternative sources of information on all sample units including non-respondents are sample frame data (to be made available as *paradata* - data about the survey data collection process) and *auxiliary data* that comes from external sources. The latter should be unambiguously linkable to all areas and/or all individuals in the sample. Then we can compare whether respondents and non-respondents differ for e.g. socio-demographic characteristics on area or individual level. Knowing about such differences allows to correct the survey estimates to some degree. But the most effective weighting corrections are only possible when there is a correlation between survey non-participation and the key survey variables (Kreuter & Olson, 2011). In conclusion, more data on the non-respondents and the non-response process should be gathered and used to improve the representativity of this, survey (like many others).

2. Evaluation of the measurement of job quality in EWCS 2010

The second pillar of quality in surveys and survey data is measurement quality (see Figure 1, Groves & Lyberg, 2010). In EWCS, measuring job quality well is essential. But in any survey and for any concept, differences between the 'true' value and the measured value can arise - the measurement errors. Good measurement instruments are needed to minimise this type of error. This starts with a good conceptualisation that needs to be turned into questions with good wording that can be understood properly by the respondents. Respondents also need to *want* to give a 'true' answer on the question, of course. Furthermore, after the questionnaires have been completed, the resulting survey variables need to be treated correctly. Six aspects of measurement quality in EWCS 2010 will be evaluated and discussed in this section.

2.1 Instrument development for EWCS 2010

We start with evaluating *how* the measurement of job quality was envisioned when the survey questions were designed. In line with the recommendations for quality control in social science research (Quality Standards Working Group, 2015), a pre-test and a thorough translation process preceded the fielding of the questionnaire for EWCS 2010 (Gallup Europe, 2013c). However, the pre-tests could not be carried out in the Candidate countries for the EU due to lack of time but were partially compensated with a bigger pilot group. Cognitive post-tests were, however, executed in the three prospective Candidate countries. Cognitive pre-testing was only done in the UK and France, with each 50 test interviews. This lead to the fine-tuning of the formulation of a few questions as well as to the exclusion of questions that were too difficult to understand or explain (see Gallup Europe, 2015f, p. 2-5).

The translation from the English base questionnaire was a multi-phased process in which research professionals were involved in each country. After the translation phase, a back-translation was ordered and reviewed by Gallup. On top of this, experts of Eurofound's European Working Conditions Observatory (EWCO) evaluated the translated questionnaires for the EU-27 countries and made the final approval. This was followed by pilot testing of all the fieldwork materials. The described procedure ascertains the quality of the instrument with regard to the wording of the questions, although the absence of pre-testing in some countries is an easy critique to make.

Whereas the reports on the fifth EWCS survey are rather detailed for the technical aspects of the questionnaire, the mechanisms behind the choice of all the included concepts, however, is less transparent. What can be deduced from the technical documents, is that experts were invited to contribute to the conceptualisation and operationalisation, namely the Advisory Committee on the European Working Conditions Survey (Gallup Europe, 2013e) but there are no details available on the decision making process for delimiting the content of the survey in the available technical documents. How and why the content of the survey has altered to a rather big degree in every edition is also hard to uncover with the publically available data. These changes lead to a multitude of differences that hamper comparison of the surveys over time.

2.2 Validity

The conceptualisation, operationalisation and scaling of core job-related concepts in EWCS 2010, such as job quality, is addressed in reports by Eurofound partners (see Eurofound, 2012; Holman, 2015; Munoz de Bustillo *et al.*, 2015; Vandenbrande *et al.*, 2013). The composition of the content of these suggested scales for job quality is theory-driven, although strongly accustomed to the available variables and items in the survey data. This clearly hints at (partial) a posteriori reasoning for the operationalisation of core concepts. Surprisingly, validity tests of the scales are mostly absent. The reports could be considered as a form of expert judgements, but concrete statistical testing of scale validity is hardly available. Only Eurofound (2012) explicitly mentions validity testing of their job quality scale, based on criterion validity. This means the operationalisation was tested and approved by its relationship with outcomes of job quality, making it a form of predictive validity testing only.

The absence of more proper validity tests for scales is rather surprising. No convergent or discriminant validity tests can be found for the scales nor do we find tests of the comparability of the dimensional structures of the scales meant for international comparative research. Some hints of differential validity can be found when studies compare sub-groups, but all in all validation of job quality scales is meagre. There is also still a clear need to develop a (gold) standard for measuring job quality in EWCS specifically and for measuring it in general as well. Theory-driven and data-driven approaches should be extensively utilised in order to come to such a highly needed standard for the scale construction of job quality for international comparisons - the main aspiration of the EWCS surveys. This would allow better cross-sectional international comparisons of job quality in Europe. Comparison over time with EWCS survey data is seriously hampered given that the content as well as the answer options of the questions have changed in each and every edition of EWCS. More stability in that regard is highly recommended for the survey.

2.3 Minimising method effects

Obtaining correct answers from respondents to the survey questions is what surveys hope for. However, deviations from the 'true' scores of survey variables are always present. People are not always consistent in the way they answer, which can remain within some regular variations. But confusion or memory effects can also play a roll, and sometimes social desirability bias, the desire to answer questions in a way that they show norm-abiding, takes control (see e.g. Groves, 1998). The latter two phenomena will create more systematic measurement errors, leading to a bias in the answers, a systematic deviation from the 'true' values.

To avoid method effects, cognitive pre-test is a minimum standard (Quality Standards Working Group RatSWD, 2015), which is mostly adhered to in EWCS 2010 (Gallup Europe, 2013c). However, in the Candidate countries this could not be done due to practical reasons, making them resort to post-tests. Another pre-emptive practice to reduce method effects is sticking with one and the same survey mode for all participating countries as different modes of data collection also lead to different kinds of measurement error (see e.g. Bautista, 2012). This already starts with preventing differences in the contact procedures between countries. Some countries participating in EWCS 2010 sent advance letters by postal mail, most did not (Gallup Europe, 2013d). In countries without address-based listings, the introduction letter was '*used by the interviewers*' (Gallup Europe, 2013c; p. 9). The use, styling and even timing of advance letters has significant effects on the (non)response processes, which can lead to a first difference between the countries participating in EWCS with regard to method effects. For example, sending the letters in advance tends to lead to better response rates in comparison to not using any at the first contact moment (e.g. Dillman *et al.*, 2009).

EWCS 2010 also allowed different data collection modes, CAPI and PAPI, probably mostly because of practical considerations such as availability of mobile computers. Both options are face-to-face surveys, interviewer-effects come into play (Bautista, 2012). Proper interviewer training should reduce such effects (Gallup Europe, 2013d, p. 31), but social desirability and satisficing

answering styles of the respondents are not always easy to detect or mediate in the survey situation itself. The workload per interviewer was limited to 20 interviews maximum, extra verification measures were taken when this number was exceeded (Gallup Europe, 2013c p. 14). About 10% of the interviews were back-checked by the local agencies. Although the report mentions amendments in case of incorrectly recorded key demographic variables, it fails to mention the incidence rate of problems and their nature in the document. As a consequence of working with interviewer-assisted data collection, the sample units and respondents are nested within interviewers, a structural element that should be taken into account with multilevel analyses to tease out the interviewer variance in the data when doing substantive studies.

2.4 Reliability

Reliability refers to the correlation between the true value and the observed value. In a perfect world, this correlation would be 1 and so would be the correlation between all the repeated measures of the instrument. In other words, the instrument would then exactly measure the same on any new occasion. As addressed in the previous sections, the survey instrument has been submitted to cognitive pre-tests and in some cases post-tests. The technical documents do not mention any issues with reliability. With regard to the reliability of the scales, the end-users should obviously include reliability coefficients when constructing composite measures.

2.5 Processing errors

Data collection and processing can be prone to human error and technical issues. Processing errors can arise while cleaning and weighting the data (Groves *et al.*, 2004). Honest mistakes can be made when programming the software and cleaning the data. In computer assisted surveys, routing problems can occur, leading to questions to be skipped or not while they should not have been. Wrong outcome codes could accidentally be assigned to variables as well. The available technical and methodological documents mention imputation and verification rules for EWCS 2010, but they are not publically available. Hence, we cannot assess whether or to what degree processing errors occurred.

2.6 Other quality characteristics

In this category, the Quality Standards Working Group (2015) emphasises efficiency, reasonableness, and up-to-datedness of the instrument. With regard to the former two quality elements, we see modest interview burden with interviews taking on average 42 minutes (Gallup Europe 2013e). The efficiency within the 2010-survey seems rather good, the efficiency over the consecutive EWCS editions, however, has room for improvement. As we already mentioned earlier, the content of EWCS has often altered over the editions. As a consequence, the comparability of the content over time is problematic. In a sense, this is a not so efficient use of a repeated survey. The use of PAPI also seems like a suboptimal in the digital age.

3. Conclusions

We assessed the fifth EWCS survey with the TSE-approach, distinguishing two major building blocks of survey quality, namely the degree of representation and measurement error. With regard to the representation, we find that the following risks and recommendations should be taken into account when designing next rounds of EWCS:

1. The target population excludes people living in institutions and in remote areas. Among both types there can be people who do fall under the definition of the inferential population. Especially the under-representation of Europeans with a paid job living in remote areas seems relevant to address given that they are chronically under-surveyed in many regards. Moreover, with current technological possibilities, alternative survey methods can be used to include these people more strongly.
2. Three different types of sample frames were allowed, each with its own typical risk of over and under-coverage, as can be seen above. The random walks are the most prone to errors and mis-representation. Only double-checking about 10% seems rather low, especially in times of Google Maps, Google Street View and other geo-spatial software being freely accessible to attempt to validate observations.
3. Sampling errors will occur to some degree in any sample, deviations from the inferential population in terms of representation of certain subgroups can (partially) be mediated by weighting the data. Mediating the sampling errors does not automatically fix non-response errors, however (see next point). Moreover, the replacement of geographical areas (the primary sampling units) in which the interview targets were hard to achieve in favour of 'easier' areas could clearly bias the realised sample. Information on neighbourhood and/or individual SES should be gathered in order to check and correct for possible SES-misrepresentations in comparison to the inferential population.
4. More paradata or auxiliary data should be included to allow to detect non-response errors and non-response bias. Better information also allows for better corrective weights.

With regard to measurement errors, there are also some issues that should be addressed in next rounds of EWCS.

1. More transparency about the decision making process for inclusion and exclusion of variables and items is needed in the technical documents.
2. Based on the available technical documents and post hoc-reports on how to measure job quality in EWCS 2010, it seems that there should be more theory-driven conceptual development of the core survey variables. It seems mostly left to the end-users to cherry-pick items to fit a scale.
3. The amount changes that is made to the content and to the answering options in each round is very problematic for the comparability over time - which, of course - should be addressed carefully anyway as these are cross-sectional surveys and not panel surveys.
4. The operationalisation of the core survey variables is also underdeveloped as no thorough validity tests of core scales, such as a job quality scale, can be found. At maximum, there is an occasional criterion validity test and the reporting of reliability coefficients. More profound validity testing, especially with regard to international comparability, should be performed and included in the reports.
5. The absence of cognitive pre-tests in some countries is an obvious remark to make.

6. The differences with regard to the use and spread of the introduction letter and the survey mode differences should be assessed and addressed.
7. There are not many details available about potential processing errors nor do the technical documents shed much light on the imputation and verification rules.
8. A more strict methodological protocol that is exactly the same for all participating countries will improve the data quality and the comparability.

All in all, the methodology behind EWCS 2010 does show a high degree of professionalism but, when assessing the survey from the ideal type TSE-perspective, there is (still) room for improvements. Because as already stated in the introduction, from the InGRID research infrastructure perspective, the EWCS data are thus a very important source to keep and to expand.

Part of the rooms for improvement, mentioned in the paper, can only be solved by more cost-consuming procedures. Financial support from other policy stakeholders to the survey could certainly help in this regard. In recent times, Eurofound has secured some national co-financing to enlarge particular national sample sizes. Other players - international science funds, European policymakers, national administrations - should also maybe step-in more.

Bibliography

AAPOR (Ed.) (2009). *Standard Definitions. Final Dispositions of Case Codes and Outcome Rates for Surveys.* Available at http://www.aapor.org/Standards-Ethics/Standard-Definitions-(1).aspx

Bautista R. (2012). An Overlooked Approach in Survey Research: Total Survey Error. In L. Gideon (ed.) *Handbook of survey methodology for the social sciences* (pp. 37-49). New York, NY: Springer.

Bethlehem J.G. (2009). *Applied survey methods: A statistical perspective (Vol. 558)*. John Wiley & Sons.

Bethlehem, J.G., & Kersten, M.P. (1985). On the treatment of non-response in sample surveys. *Journal of Official Statistics, 1*(3), 287.

Dillman, D.A., Smyth, J.D., & Christian, L.M. (2009). *Internet, mail, and mixed-mode surveys. The tailored design method*. Wiley: New Jersey.

Eurofound (2014). *Policy lessons from the fifth EWCS: The pursuit of more and better jobs*. Luxembourg: Publications Office of the European Union.

Gallup Europe (2013a). *5th European Working Conditions Survey, 2010. Technical Report*. European Foundation for the Improvement of Living and Working Conditions.

Gallup Europe (2013b). *5th European Working Conditions Survey, 2010. Sampling Implementation Report*. European Foundation for the Improvement of Living and Working Conditions.

Gallup Europe (2013c). *5th European Working Conditions Survey, 2010. Quality Assurance Report*. European Foundation for the Improvement of Living and Working Conditions.

Gallup Europe (2013d). *5th European Working Conditions Survey, 2010. Technical Report*. European Foundation for the Improvement of Living and Working Conditions.

Gallup Europe (2013e). *5th European Working Conditions Survey, 2010. Quality Assessment of the 5th European Working Conditions Survey*. European Foundation for the Improvement of Living and Working Conditions.

Gallup Europe (2013f). *5th European Working Conditions Survey, 2010. Pre-test Report*. European Foundation for the Improvement of Living and Working Conditions.

Groves, R.M. (2006). Non-response rates and non-response bias in household surveys. *Public opinion quarterly, 70*(5), 646-675.

Groves, R.M. & Couper, M.P. (1998). *Non-response in household interview surveys*. New York: Wiley.

Groves, R.M, Fowler, F.J., Couper, M.P., Lepkowski, J.M., Singer, E., & Tourangeau, R. (2004). *Survey methodology*. New Jersey: Wiley.

Groves, R.M. & Lyberg, L. (2010). Total Survey Error: Past, present and future. *Public Opinion Quarterly, 74*, 849-879.

Kreuter, F., & Olson, K. (2011). *Multiple auxiliary variables in non-response adjustment*. Sociological Methods & Research, 0049124111400042.

Lynn P. (2003). PEDASKI: Methodology for Collecting Data about Survey Non-Respondents. *Quantity and Quality, 37*, 239-261.

Quality Standards Working Group (2015). *Quality Standards for the Development, Application, and Evaluation of Measurement Instruments in Social Science Survey Research* (RatSWD Working Paper Series 245). Berlin: Federal Ministry of Education and Research.

InGRID
Inclusive Growth Research Infrastructure Diffusion

Referring to the EU2020-ambition of Inclusive Growth, the general objectives of InGRID – Inclusive Growth Research Infrastructure Diffusion – are to integrate and to innovate existing, but distributed European social sciences research infrastructures on 'Poverty and Living Conditions' and 'Working Conditions and Vulnerability' by providing transnational data access, organising mutual knowledge exchange activities and improving methods and tools for comparative research. This integration will provide the related European scientific community with new and better opportunities to fulfil its key role in the development of evidence-based European policies for Inclusive Growth. In this regard specific attention is paid to a better measurement of related state policies, to high-performance statistical quality management, and to dissemination/outreach activities with the broader stakeholder community-of-interest, including European politics, civil society and statistical system.

InGRID is supported by the European Union's Seventh Programme for Research, Technological Development and Demonstration under Grant Agreement No 312691.

More detailed information is available on the website: www.inclusivegrowth.be

Co-ordinator
Guy Van Gyes
Monique Ramioul

KU LEUVEN HIVA
RESEARCH INSTITUTE FOR WORK AND SOCIETY

Partners
TÁRKI Social Research Institute Inc. (HU)
Amsterdam Institute for Advanced labour Studies, Universiteit van Amsterdam (NL)
The Swedish Institute for Social Research, Stockholms Universitet (SE)
Fachbereich IV, Wirtschafts- und Sozialstatistik, Universität Trier (DE)
Centre d'Estudis Demogràfics, Campus de la Universitat Autònoma de Barcelona (ES)
Luxembourg Institute of Socio-Economic Research (LU)
Centre for Social Policy, Universiteit Antwerpen (BE)
Institute for Social & Economic Research, University of Essex (UK)
Bremen International Graduate School of Social Sciences, Universität Bremen (DE)
Centre d'Etudes de l'Emploi et du Travail (FR)
The Centre for European Policy Studies (BE)
Dipartimento di Economica e Menagement, Università di Pisa (IT)
Social Statistics Division, University of Southampton (UK)
Luxembourg Income Study, asbl (LU)
WageIndicator Foundation (NL)
School of Social Sciences, The University of Manchester (UK)

InGRID

Inclusive Growth Research
Infrastructure Diffusion
Contract No 312691

For further information about the
InGRID project, please contact
inclusive.growth@kuleuven.be
www.inclusivegrowth.be
p/a HIVA – Research Institute
for Work and Society
Parkstraat 47 box 5300
3000 Leuven
Belgium