Methodological note

Minimising the impact of COVID-19 on data quality in the agricultural census

EUROSTAT, DIRECTORATE E

UNIT E1 — AGRICULTURE AND FISHERIES

08 JULY 2021

Acknowledgements

The European Commission expresses its gratitude and appreciation for the preparation of the report to the following members of the ESS ([1]) discussion group:

Experts from national statistical institutes:

- Pedro Campos (Statistics Portugal)
- Cathal Doherty (Central Statistics Office of Ireland)
- Karsten Larsen (Statistics Denmark)
- Ágnes Patay (Hungarian Central Statistical Office)

Expert from the Food and Agriculture Organization of the United Nations (FAO):

- Jairo Castaño

Coordinators from Eurostat:

- Denisa Camelia Florescu
- Helena Ramos

The European Commission also thanks the following experts from the national statistical institutes for their contributions:

- María Cimadevilla (Spanish National Statistical Institute)
- Predrag Cvjetićanin (Croatian Bureau of Statistics)
- Aušra Jablonskienė (Statistics Lithuania)
- Ilze Januška (Central Statistical Bureau of Latvia)

([1]) European Statistical System

Contents

1. Introduction

In several EU countries, the COVID-19 pandemic had a big impact on the census, mainly by causing delays. The delays affected several activities, especially printing questionnaires, training interviewers and supervisors, contracting, data collection and access to administrative registers.

Despite the difficulties, the reference year 2020 was kept for all countries, except Portugal which had 2019 as reference year.

In response, the European Statistical System working group on ‘Structure of Agricultural Holdings’ set up a small task force to discuss the practices used in Europe to minimise these difficulties. This report presents the best of these practices.

Although they apply to the COVID-19 pandemic, these practices are also applicable in any other similar situation of general restrictions, and even in normal situations.

2. General situation in EU countries

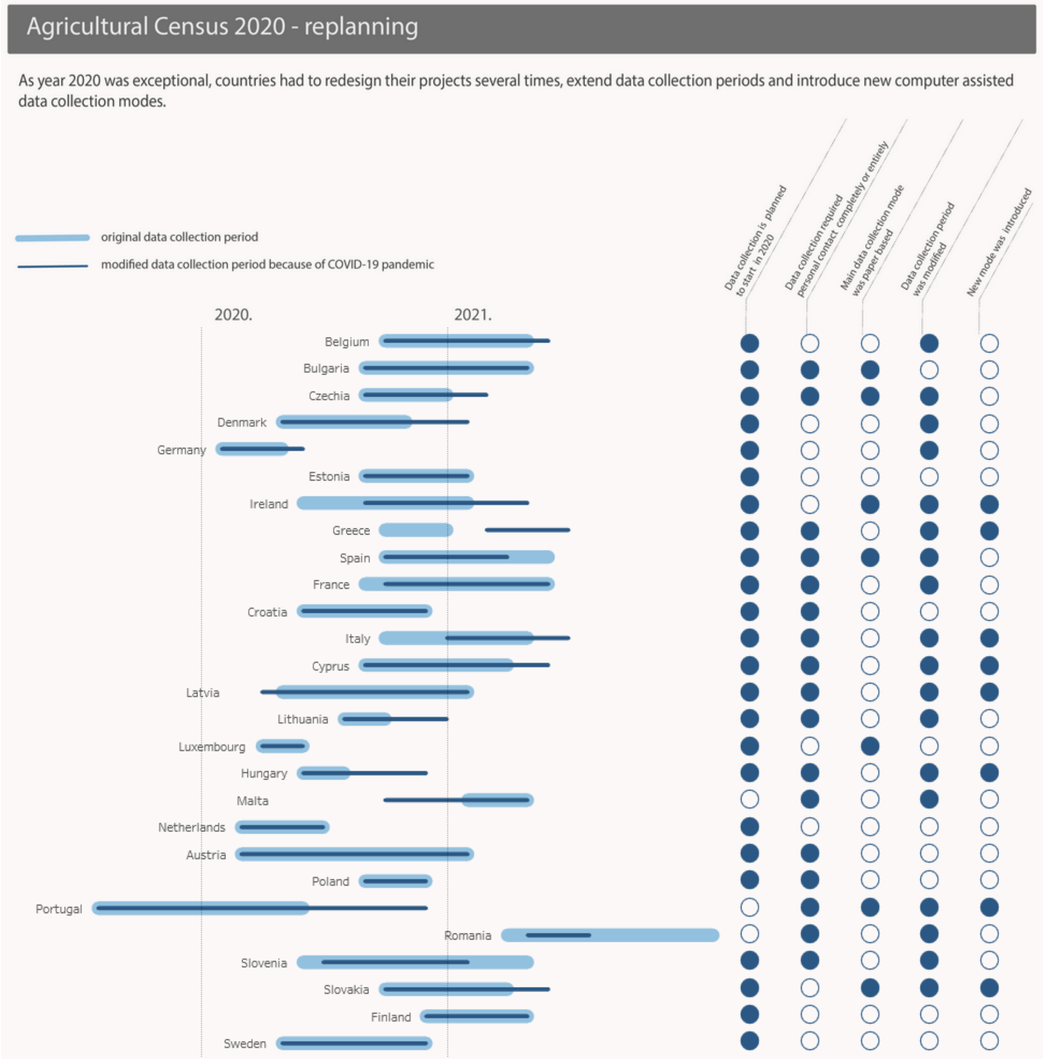
While some countries suffered no impact of COVID-19 pandemic on their agricultural census, others were significantly affected.

The countries that suffered no or low impact had had earlier improvements in their national statistical systems, a wide range of IT solutions and regular use of data from administrative registers. These factors enabled them to significantly reduce their reliance on physical contact for tasks such as final preparation of field activities, training and data collection.

Most countries collected data for the 2020 census at least partly in electronic form. Still, two thirds of them had to move or extend the data collection period. This was usually accompanied by the introduction of a new data collection mode.

The most affected countries were those where at least part of the data were collected through face-to-face interviews. To collect data from all farmers, 30 % of countries introduced some kind of additional computer-assisted data collection mode.

Figure 1: Changes in data collection periods and modes caused by the COVID-19 pandemic



Although the data collection period varied across countries, the reference year 2020 ([2]) remained unchanged, to comply with the Regulation (EU) 2018/1091.

The pandemic also forced countries to change their census communications and outreach. They had to adapt their publicity and information campaigns to encourage people to use CAWI (Computer-Assisted Web Interview), CATI (Computer-Assisted Telephone Interview) and POST (data collection by post) to ensure social distancing.

It was also more difficult to get respondents' attention. Fortunately, farmers could be reached via numerous organisations and agricultural chambers that were able to help the national statistical bodies promote the data collection.

([2]) Only Portugal carried out the agricultural census with the reference year 2019. Regulation (EU) 2018/1091 granted Portugal a derogation, to avoid an overlap with the population and housing census.

3. Quality aspects affected by the pandemic

Some countries assessed the overall impact of the pandemic on the census data quality to be minor. In these countries, the pandemic had not caused errors other than the usual refusals to provide data, which were negligible.

In other countries, the pandemic affected various quality aspects, but the overall impact could not be quantified, at least in the initial stages.

3.1 Relevance

Generally speaking, as 2020 was not a normal year because of the pandemic, the census data might not reflect the reality of the agricultural sector. This also happens in years affected by extreme weather conditions.

However, the variables collected in the census refer to the structure of farms and not to production figures for crops and livestock (more likely to be affected by the pandemic). Only some slight changes are expected in the number of farms in certain agricultural sectors and in variables related to labour force and other gainful activities.

Therefore, the overall impact of the pandemic on the relevance of the census data is generally pre-assessed as not significant.

3.2 Accuracy

The pandemic is likely to increase sampling errors, coverage errors, non-response errors, measurement errors, processing errors and model assumption errors.

- **Sampling errors**

Farms were more reluctant to participate. It is known that unit non-response decreases the net sample size. When unit non-response is missing at random, the group(s) of respondents can be assimilated to random sample(s) drawn from the population (or from strata of the population).

Unit non-response is addressed by imputation or by re-weighting the data collected from respondents. Re-weighting (increasing respondents' weights) increases sampling errors.

- **Coverage errors**

Over-coverage and under-coverage: The pandemic affected some holdings operating in certain agricultural sectors and would slightly make sampling frames outdated. Outdated sampling frames introduce bias in samples for subsequent agricultural surveys.

Multiple listings: Because of the pandemic, face-to-face interviews were postponed. Data collection using CAPI (Computer-Assisted Personal Interview) or PAPI (Paper and Pencil Interview) was usually conducted later than CAWI, CATI and POST. Some farmers did not remember they had already answered the survey and they or their family members answered a second time. This resulted in duplicated farms.

- **Non-response errors**

Non-response errors refer to unit non-response, item non-response or values flagged as erroneous during editing and therefore treated as non-response.

Some farmers were reluctant to accept CAPI and PAPI interviews during the pandemic and did not respond. It is assessed that the pandemic had only a slight effect on unit non-response, as countries managed to collect the data from most of the respondents.

Data on complex variables collected using CATI led to some values being found to be implausible and treated as item non-response.

- **Measurement errors**

Postponements of CAPI and PAPI data collections made it harder for farmers to clearly recall and accurately report the data for the same, fixed reference period.

Data on complex variables collected using CATI led to some values that were a rough approximation of the real values, not detected as implausible and used to estimate the aggregates.

- **Processing errors**

Unit and item non-response is in many cases addressed by imputation. Imputation is a source of processing errors.

- **Model assumption errors**

Imputation models (used to cover for non-response) or calibration models (used to improve the quality of the final data) rely on assumptions. They add an additional layer of complexity, especially when addressing the accuracy of the process, since the validity of the model is also a source of uncertainty.

3.3 Timeliness and punctuality

Activities, especially data collection, were postponed or prolonged. Where countries could not compensate for the time loss, this delay caused an unwanted shift in the national publication of data and its transmission to Eurostat.

3.4 Coherence and comparability

Postponing the census could have caused unwanted shifts in reference periods across countries and jumps in the time series. Despite the difficulties, the 2020 reference year was kept by all countries except Portugal, which had 2019 as reference year (under a derogation from Regulation (EU) 2018/1091). Therefore, coherence and comparability are generally assessed as not affected.

While these are brief general quality assessments, the effect of the pandemic on data quality can be better assessed and known after the analysis of data is finalised in all countries.

4. Best practices for minimising the impact of pandemics

This chapter groups the best practices in two categories: actions to prevent a loss of quality in the collected data in the early stages of census organisation and actions to remedy quality loss after it occurs.

4.1. Mitigating loss of quality

Preliminary evidence shows that the countries who fared better than others during the pandemic used multiple remote data collection methods in parallel (e.g. CAWI, CATI, post) and sourced census data from administrative registers (Castano, 2020).

Some of them used e-learning solutions, and widened the use of remote methods, at the expense of face-to-face interviews, to comply with increasing social distancing regulations.

This strategy worked more effectively when:

- there were several devoted applications and clearly defined teams for each data collection method;
- the IT infrastructure of the census statistical body allowed steady remote access to systems, so the census work could continue through remote working;
- training could continue via e-learning;
- usual office processing tasks were adapted so they could be performed remotely;
- publicity and information campaigns were adapted in good time to encourage agricultural producers to use alternative data collection methods;
- call centres for Q&A (hotlines) and a website with FAQs were set up.

The sections below list additional practices recommended for mitigating the impact of the pandemic on data quality, presented by subject matter.

4.1.1 COMMUNICATION

Despite the difficulties, the best possible efforts should be made to gather survey responses. In this respect, the communication side toward the observation units (the farmers) is very important. Many countries in Europe, as well as the U.S. Department of Agriculture (USDA) publish advertisements in trade journals and pre-interview letters are sent to sample members. These are seen as methods for reducing unit non-response.

The following best practices could be mentioned:

- Lobby local administration units and farmers' associations to build support and promote the census among the farmers;
- Include alerts on the websites of the statistical offices, in the reporting tools for businesses, etc.;
- Persuade especially the most influential holdings to respond;
- Emphasise to farmers that the census is mandatory;
- Organise a small lottery (with small prizes) for the respondents and communicate that in advance to the farmers.

4.1.2 DATA COLLECTION

A total census in the agricultural sector means that each and every farmer completes the questionnaire, or at least all farms above a certain size. But sometimes this is not possible in practice. There will always be some kind of non-response.

We must be realistic and realise that there are interests on both sides (statisticians and farmers). It is in our interest that the questionnaires are completed. Farmers might not feel that it is also in his own interest – they often even see statistics as an unnecessary burden. With this in mind, we should do whatever we can to make the contact between the farmer and the statistical office as easy as possible.

The following advice could be given:

- Pre-fill answers in questionnaires to be checked by farmers;
- Prepare explanatory notes on the questionnaires for enumerators and respondents;
- Pre-test questionnaires (e.g. expert reviews, cognitive interviews, advance field testing);
- Accept proxy interviews when the farmer is a person at risk (however, there is likely a trade-off between non-response errors and measurement errors);
- Extend the data collection period;
- Shorten questionnaires to request only regulatory information;
- Provide the farmers with both an e-mail address and a phone number so they can easily contact the statistical offices;
- Even if the standard response format is an online questionnaire, let farmers have the option of giving information by telephone interview;
- Avoid sending a farmer a paper questionnaire by post. These are very often filled in wrongly or incompletely (leading to extra work re-contacting the farmer or trying to guess missing information).
- In addition to a standard procedure with written reminders for the farmers, telephone reminders should be used. And don't be satisfied with a promise from the farmer to 'send the questionnaire soon' – try to get the information on the spot, during the same call.
- (Re)contact farmers at different times and on different days, carefully planned through a call scheduling system.
- In cases (like Portugal) where most contacts are made through face-to-face mode using CAPI, this makes it possible to correct any errors and non-responses directly with the respondents.
- For total (unit) non-response, reorganise your data collection infrastructure by reallocating good resources (interviewers) to places where they are needed most.

4.1.3 RECALL ISSUES

The pandemic forced some adjustments, such as postponing the start of the collection work or extending the collection period. These adjustments may make it more difficult for farmers to recall information as the census reference period remains the same.

To counter this:

- Add some introductory material that prepares the farmer for the question. This allows them to gather their thoughts before suppling a response;
- Ask multiple questions. This improves the probability that an event will be recalled and reported;
- When long recall periods have to be used, use a life events calendar to improve reporting

(e.g. associate the periods with certain important moments regarding political decisions on the pandemic that were made in that country).

- When possible, pre-fill answers in difficult-to-recall questions, to help farmers respond.

4.1.4 MODULE DATA

4.1.4.1. Complex variables

The increased use of remote data collection methods, such as CATI, CAWI and POST poses some problems in administering the census modules ‘Labour force and other gainful activities’ and ‘Animal housing and manure management’.

These modules involve complex variables that present challenges for respondent's ability to both understand and report accurately. Examples of these variables include working hours of the labour force, family members involved in farm work and other gainful activities, and manure management.

- Formulate complex questions in clear simple language, wherever possible using vocabulary familiar to the farmers, so all farmers understand the concepts in the same way;
- Separate questions considered too complex into multiple questions. However, since this increases the total number of questions, there is likely a trade-off between simplification and an increased burden on respondents;
- When relevant, provide definitions before the question itself is asked;
- Make clear to farmers that being accurate is very important for the census;
- Make sure the reference period for the questions is unambiguous;
- Increase the gross sample size, to alleviate the potential negative impact on unit response rate, especially in modules with complex variables.

4.1.4.2. Sensitive variables

The module on manure management has some sensitive variables that require answers that reflect actual practices and not what is socially desirable. The use of self-administrative questionnaires, such as in CAWI and POST, may facilitate factual answers from farmers.

- The increase in the share of data collected by CAWI and POST during the pandemic is likely to favour more accurate answers on sensitive variables;
- As far as possible, rephrase sensitive questions so they look less threatening;
- When the interview is face-to-face (e.g. CAPI), the tablet can be handed to farmers to answer sensitive questions themselves, without the interviewer’s involvement;
- Emphasise the importance of accurate answers on manure management data.

4.2. Addressing loss of quality

In some cases, the reduction in the quality of collected data could not be prevented. In these cases, the national statistical bodies took several steps to assess and address it. These include dealing with missing values, outliers, duplicates and other errors.

This chapter presents best practices, grouped by the order in which they are applied.

4.2.1. IDENTIFYING PROBLEMS IN DATA AND CALLING BACK FARMERS OR USING ADMINISTRATIVE DATA TO COMPLETE OR CORRECT THE MICRO-DATA

The data quality risks being lower than in normal circumstances due to the lower response rate and timeliness issues influenced by the pandemic. Some consequences of these aspects are:

- **missing values**

  – total (unit non-response): answers for the whole farm are missing;
  – partial (item non-response): answers for some variables in the census are missing.

- **data outliers**

  – values that are significantly different from other data points, potentially causing problems in statistical procedures.

- **duplicates** e.g.
  – members of family farms reporting the same administrative identification numbers
  – members of family farms reporting the same data on the same address
  – private farms reporting the administrative identification number of an agricultural enterprise.
- **other types of errors**

  – systematic errors
  – random errors.

Best practices to deal with these errors consist of:

- using validation rules embedded in the data collection applications. Some countries created additional validation rules to detect inconsistencies during the period affected by the pandemic;
- identifying outliers through the use of statistical models (Blázquez-Garcia et al, 2021):
  - extreme outliers (Median of the Absolute Deviations (MAD) – MAD-score (base and stats); composed-MAD – MAD-score (base and stats); Isolation Forest – tree depth (isotree));
  - soft outliers: iterative ARIMA model with outliers – t-stat (tsoutlier); STL-decompose with remainders – IQR-score (anomalise).
- analysis of standard deviations within strata;
- comparison of micro-data to administrative sources and control surveys using matching keys such as address and identification number.

To complete the missing answers or replace the erroneous values of the questionnaire, the main initial strategies are recontacting the farmers and using administrative data.

4.2.2. RECORDING IN THE DATABASE THE REMAINING MISSING AND INVALID VALUES FOR EACH VARIABLE FOR EACH FARM

Missing values correspond to the absence of data items for a subject. This may be crucial, since these data items may hide some information that may be important. Missing data pervasively exist in most real-world datasets. Identifying the amount of non-responses is important for performing the estimation process and avoiding bias. Bias can be calculated as the product of two components: non-response rate and the difference between observed and non-respondent answers.

In the process of data preparation, statistical offices need to keep track of the remaining missing values and invalid values. This allows them to identify the (item) non-respondents in the dataset, trace them back after the imputation is done and calculate (item) non-response rates.

Usually problems with missing data can be of two kinds: total or partial non-response:

- unit non-response (falling under total non-response) – when a farm fails to respond to the survey invitation.
- item (or partial) non-response – when a farm fails to complete an item on the questionnaire.

4.2.3. CONDUCTING A NON-RESPONSE ANALYSIS

Non-response leads to an increase in variance due to a reduction in the size of the sample. If the non-respondents have characteristics of interest that are different from those of the respondents, non-response can produce bias, as mentioned above. The sampling error may increase if imputed data are treated as though they were observed data.

In the agricultural census, it is possible that total non-response persists, although this number is very small (about 1 % of the total universe in Portugal).

For partial non-response, in Statistics Denmark most likely the most common method used has been that the interviewers have completed the questionnaire themselves, using available information. It could for instance be last year’s questionnaire. However, this solution is not satisfactory because the content of farm structure surveys is not identical from survey to survey. A more valid solution is the use of donor imputation: farmer A has completed the questionnaire but B has not. If we have reasons to believe that A and B look like each other, we can transfer A’s survey information to B. We will then provide data for the non-responding farms which of course is not authentic for each farm but which most likely on average will come close to the truth. We check whether A and B look like each other by comparing respondents and non-respondents on variables that non-respondents might have answered in the current rounds. Such analysis should assess whether the respondents are in any way different from the non-respondents.

In general, there are several ways to conduct a non-response analysis:

- compare respondents and non-respondents on auxiliary variables available in the sampling frame or administrative sources. The auxiliary variables should be (strongly) correlated with the missing variables collected in the census.
- compare respondents and non-respondents on the variables collected in previous rounds, whose values are missing in the census;
- compare respondents and non-respondents on variables that non-respondents might have answered in the current rounds (as mentioned above). These variables should be strongly correlated with the missing variables collected in the census.
- take a sample of non-respondents, pursue them to respond and compare their answers to the respondents' answers;

- compare answers provided by early respondents (no reminder sent to them) and late respondents (one or more reminders sent to them).

The analysis will tell whether the non-respondents e.g. are big or small farms, have young or old managers, have systematically low or high values for the main variables, etc., compared to the respondents.

The analysis allows us to identify the non-response mechanism:

- if the non-respondents and respondents are similar, then the non-response is missing at random and, once imputed or re-weighted, the quality of the estimates is maintained.
- if however the comparison shows that non-respondents are different from respondents, then the non-response is not missing at random and the estimates are subject to bias. The imputation algorithms must be very reliable.

Statistical bodies can construct non-response models using information gained in initial survey rounds to model panel attrition in later waves. Some studies in social and agricultural surveys have used regression models with several classes of variables as predictors, related both to the 'contactability' of the household / holding and the propensity to cooperate, once contacted (McCarthy, J., Jacob, T., and McCracken, A., 2010).

4.2.4. IMPUTING AND CALIBRATING THE DATA

The problem of missing values – for which an imputation process is needed – occurs both for data collected in traditional surveys and for administrative data. Statisticians often create a complete dataset prior to the estimation stage, by replacing the missing values with estimated values from the available data. This process is referred to as imputation.

To impute the missing values in a dataset, several methods are available (Eurostat, 2017), and each may be useful in different contexts.

Possible imputation methods include:

- deductive imputation
- model-based imputation (including mean, ratio, and regression imputation)
- donor imputation (including cold deck, random hot deck, and nearest-neighbour imputation, as well as predictive mean matching).

The objective in donor imputation is to fill in the missing values for a given unit by copying observed values for another unit, the donor. Typically, the donor is chosen in such a way that it resembles the imputed unit as much as possible on one or more background characteristics. The reason for this is that if the two units match (exactly or approximately) on a number of relevant auxiliary variables, it is likely that their scores on the target variable will also be similar.

The standard procedure of imputation requires us to know something about the donors. In agricultural surveys, if we know nothing about donor B apart from that the farm exists, then the donor could be chosen completely randomly.

But most often we will know something. It could for instance be information from business registers, livestock registers, IACS (Integrated Administration and Control System), tax registers or cadastre information, giving us an idea of items like the farm's area size and its economic size. Using such available information when selecting the donor is certainly far better than random selection.

Some recommendations to bear in mind when doing imputation:

- Avoid mean imputation, because it decreases variance.
- Use donor imputation where, for each non-response farm, select a farm that has completed the questionnaire. An example on farmer's education: we know that the age of the farmer, the farm size and the gender of the farmer are related to education. Small farmers are more likely to have 'practical experience only'. The same is true for old farmers and women. So for a farmer to be imputed we find a donor with the same gender, almost the same size agricultural area and almost the same age. But for other parts of the questionnaire, for instance animal housing, other relations might exist, meaning that we should find a new donor and not use the same donor for all parts of the questionnaire.
- Favour stochastic imputation methods to account for some of the variance of the non-respondents. However, even after adding random disturbances to the imputations, the variance of an estimator can still be underestimated, because the uncertainty of the imputation model itself is not taken into account. Therefore, even better, use multiple imputation. This enables you to obtain valid estimates of standard errors and confidence intervals.
- Apply imputation methods in imputation cells created by the intersection of auxiliary variables (stratification variables or variables used as publication domains). The auxiliary variables have to strongly correlate to the variables to be imputed. This makes the variables to be imputed as homogeneous as possible within the cells, and the response mechanism MAR (missing at random) can be assumed within the cells. The choice of imputation cells corresponds to an implicit model for the variables to be imputed.
- Use the best non-missing auxiliary variables to predict the values of the missing variables. The auxiliary variables have to strongly correlate to the variables to be imputed. This determines the quality of the imputation.
- Use calibration techniques to adjust the final dataset to correspond to frame population and reliable administrative sources.

What is the maximum acceptable imputation rate?

The quality of the results depends not only on the amount of missing data but also on:

- **the extent to which the characteristics of respondents differ from the characteristics of non-respondents.**

  If respondents are different from non-respondents then it does not matter if only 5 % or 10 % of the data is missing, the estimates are biased in the absence of proper imputation algorithms.
- **the quality of the imputation algorithms.**

  If the imputation method is poor – i.e. it predicts the missing values in a biased manner – then it does not matter if only 5 % or 10 % of the data is missing (it will still produce biased results even if maybe tolerably so). The larger the share of missing data, the greater the importance that the imputation algorithm is valid.

Denmark anticipates an imputation rate of about 10 %. Most cases imputed are related to situations that occur in a data collection in an usual context. The average non-response farm is smaller than other farms and it means that farms with authentic information will possess a good deal more than 90 % of the agricultural area and livestock. It should also be noted that some of the imputed information comes from IACS (for crops) and the livestock register (for cattle).

4.2.5. ASSESSING THE RESULTS AFTER IMPUTATION

Finally, once missing and erroneous values are imputed, the following best practices are relevant to mention:

- Analyse the results with and without the imputed data, with both tables and graphs. This analysis shows the impact of the imputation on the final results.
    - If that impact is high, make sure the imputation model is very reliable.
    - If imputation based on models was used, check the distribution of imputed values for possible inappropriate or outlying values.
    - If real donors were used, check that donors have been used in adequate numbers.
- Analyse the difference between the variance calculated assuming no imputation and the variance incorporating the specific variance arising from imputation.

  This analysis measures the additional variability caused by imputation.

  The methods to estimate the variance accounting for imputation are usually multiple imputation and resampling techniques (Eurostat, 2013).
- Compare the aggregated, imputed final results against reliable sources and control (post-enumeration) surveys, at overall level and by several dimensions. This analysis measures the overall bias, including the one caused by imputation. Reliable sources can be administrative sources or other data collections.
- Measure, document and publish item and unit non-response rates as well as relative standard errors. The non-response rates can be unweighted, design-weighted and size-weighted (Eurostat, 2020).

5. Positive effects of the pandemic on quality & future data collections

While the pandemic put extra pressure on national authorities, a number of positive outcomes have been realised during data collection. These outcomes affected the data quality and can be observed within the new working environment, across data collection strategies, and will potentially be observed in future data collection strategies.

Countries used remote working extensively, creating a new work environment, training moved online and many processing tasks were carried out by teams remotely.

Many countries were forced to modernise and innovate just before or during data collection, leading to more variety in collection methods and a higher likelihood of countries using multiple modes, thus reducing the burden on the respondent.

Also, there is now an opportunity to modernise future data collections by adapting the mode of collection to suit the respondent. Several countries now have a rich data source that describes the respondent and the mode that was used to complete the 2020 agricultural census; this source of information can be used to target respondents with a particular mode.

5.1. Working environment

For staff working on the census, most tasks moved online, so teams are now familiar with the available online tools and have seen how useful they were during planning, data collection and data processing.

It was easier to train larger numbers of interviewers directly using an online platform, compared with face-to-face training. Interviewers also had the possibility to return to recorded versions of training, which enabled follow-up questions and increased familiarity with content.

Many office tasks also moved online, including data processing, which in many cases was found to be very efficient. It is likely that many of the remote working tasks that were once performed face-to-face will remain as remote tasks, due to their increased efficiency and ease of access. There are benefits to both on-site and remote working, so a type of blended environment may replace the fully on-site work of the past.

5.2. Data collection

A major result of the pandemic has been the introduction of new data collection modes across countries. Even countries that planned paper-based data collection introduced new modes such as CAWI or CATI.

The increased use of CAWI was particularly beneficial for statistical bodies and respondents; statistical bodies now had more possibilities for responses and respondents had less burden when responding. It is likely that increasing the number of modes has led to higher response rates, not to mention improving respondents' impression of the statistical bodies.

Respondents have adjusted their behaviour, and in some countries a higher than estimated preference for CAWI response has been observed. Respondents have also registered with administrative entities, both increasing coverage and making it more likely that a respondent remembered their unique registration number. These factors increased the quality, when survey and administrative data are linked.

5.3. Future

Multiple data collection modes have been implemented during the 2020 agricultural census. It is clear that all countries had to adjust their data collection strategy in some way due to the pandemic.

Countries that used a mode they had not planned to use now have an opportunity to modernise future data collections. Targets can be set over the coming years to move from paper-based to web-based collection where possible. Data collection strategies of the future should be adapted to suit respondents, which will increase the chances of a high response rate and lead to innovative methodologies for data collection.

A rich data source should be available in most countries detailing the response mode linked with several characteristics of the respondent. When choosing samples in future, it should be clear which respondents used a particular mode during the census, and these respondents could be targeted with the same mode.

Data can also be analysed to determine if particular respondent profiles exist across modes. For example, early analysis in Ireland has shown that age has an impact on mode, with CAWI respondents 10 years younger on average than CATI respondents, and the average age of postal respondents typically lying between those who respond via CAWI or CATI. There also appears to be a regional influence on the mode used, with rural locations more likely to use CATI and Postal than CAWI. This analysis will be extended to examine the quality of responses across modes and item response rates across modes, as well as incorporating other characteristics such as farm type and farm size.

Future data collections in agriculture can build on the modes used in the 2020 agricultural census, thus extending the possibility for increased response rates and leading to modern, innovative data collection strategies.

6. Conclusions

EU countries applied and recommend a wide range of measures to minimise the effect of the pandemic on the quality of the statistics of the agricultural census. Those actions go from action in the planning stage to preserve the quality of the data to be collected as far as possible, to steps taken in the analysis stage to redress the quality reduction, once that happens.

The key success factors in adapting to the new circumstances are the extensive use of administrative data and investment in stable and reliable IT infrastructures for census organisation and remote connection.

The experience of the COVID-19 pandemic validated such developments and further motivates the national statistical organisations to speed up the pace of innovation.

7. References

- Blázquez-Garcia et al (2021). *A Review on Ofloredautlier/Anomaly Detection in Time Series Data.* ACM Computing Surveys, Volume 54, Issue 3, June 2021
- Castano, J. (2020). *Censuses of agriculture and COVID-19: Global situation and lessons*. Statistical Journal of the IAOS. 36: 861–865
- Castano J. (2020*). Impact of COVID-19 on national censuses of agriculture (Status overview).* FAO Policy Brief, Rome
- Castano J. (2020). *Impact on censuses of agriculture and some mitigation measures*. In: *How COVID-19 is changing the world: a statistical perspective, Volume II*, Committee for the Coordination of Statistical Activities (CCSA).
- De Leeuw, E. D., Hox, J. J., and Dillman, D. A. (2008). *International handbook of survey methodology.* Taylor & Francis Group/Lawrence Erlbaum Associates
- EUROSTAT (2021). *ESS Vision 2020 ADMIN (Administrative data sources)*
- EUROSTAT (2020). *European Statistical System handbook for quality and metadata reports*
- EUROSTAT (2017). *Handbook on Methodology of Modern Business Statistics*
- EUROSTAT (2013). *Handbook on precision requirements and variance estimation for ESS households surveys.* Methodologies & Working papers, Publications Office of the European Union, Luxembourg
- Floyd J. Fowler, Jr., Floyd J.. Fowler (1995). *Improving Survey Questions: Design and Evaluation.* Applied Social Research Methods. Publisher SAGE
- Groves (2009). *Survey Methodology*. Wiley Series in Survey Methodology. 2nd Edition. Paperback
- GOPA (2020). *Mission report*. 27–28 October 2020, online meeting with the staff of the Statistical Office of Montenegro
- McCarthy, J., Jacob, T., and McCracken, A. (2010). *Modeling Non-response in National Agricultural Statistics Service (NASS). Surveys Using Classification Trees*, USDA
- Statistics Finland (2007). *Quality Guidelines for Official Statistics*, 2nd revised edition