



Food and Agriculture  
Organization of the  
United Nations



# Regional Workshop on Volume 2 “Operational guidelines” of the FAO World Programme for the Census of Agriculture 2020 (WCA 2020)

Amman, Jordan, 1-4 April 2019

## Reconciliation of current statistics with census results

*Technical Session 8.4*

**Oleg CARA**  
Statistician  
Agricultural Census and Survey Team  
FAO Statistics Division



# Contents

- 1. Introduction**
- 2. Objectives of reconciliation**
- 3. Common sources of discrepancies between Census and Surveys data**
- 4. Correcting discrepancies**
  - Selecting the key variables for reconciliation
  - Computing the gaps
  - Making the diagnosis
  - Corrections
- 5. Methods used for data reconciliation**
- 6. Country example**



# 1. Introduction

- Agricultural census is main source of agricultural statistics which is usually conducted every five or ten years to provide structural agricultural and benchmark data, but also frame for subsequent annual surveys.
- However, a problem affecting survey data is the growing time gap with the census and the resulting obsolescence of the census frame. There is a need for survey data calibration and sample frame updating.
- When a new census is conducted, a new frame becomes available and subsequent surveys are based on these new information. This could generate discrepancies between two datasets (before and after the new census).

# 1. Introduction (Cont'd.)

- These discrepancies may originate from various sources and a number of solutions exist to address them. Reconciliation consists in comparing the estimates from previous surveys and the new census results and correcting any discrepancies between them.
- The main objective of reconciliation is to improve survey estimates using new census data and to draw lessons for future surveys. Reconciliation may also help to correct some census data considering survey estimates.
- Similarly, in countries where current statistics come from expert opinion or other sources different to surveys, reconciliation helps to improve those statistics using the census data and to calibrate the methods used.
- In some countries, data reconciliation is part of the revision policy of the statistical agency (e.g. USA and Canada).

## 2. Objective of reconciliation



- The following are the main objectives of the reconciliation of censuses and surveys data (Bernhardt and Helfand, 1980):
  1. To measure the extent of differences between censuses and current surveys in coverage, classification and data;
  2. to determine why the differences occurred;
  3. To identify systematic errors, make corrections during census and survey processing, and take steps to minimize similar future errors;
  4. To identify random errors, including reporting errors, and make corrections, thus improving the annual and census estimates;
  5. To improve coverage in survey through the addition of new units;
  6. To improve the quality of annual and other current statistics using census data for benchmarking;
  7. To serve as a guide in planning for future surveys and censuses.

### 3. Common sources of discrepancies between Census and Surveys data



- Some gaps between surveys estimates and census data may be normal due to many factors such as different implementation time.
- However, other gaps may arise from both sampling and non-sampling errors.
- Types of errors found in census work include sampling errors (which occur when sampling is used) and non-sampling errors (the result of mistakes committed in various phases of the census work).
- The focus here is on the discrepancies caused by non-sampling errors.
- Technical session 3 discussed the main non-sampling errors, comprising coverage and content errors.

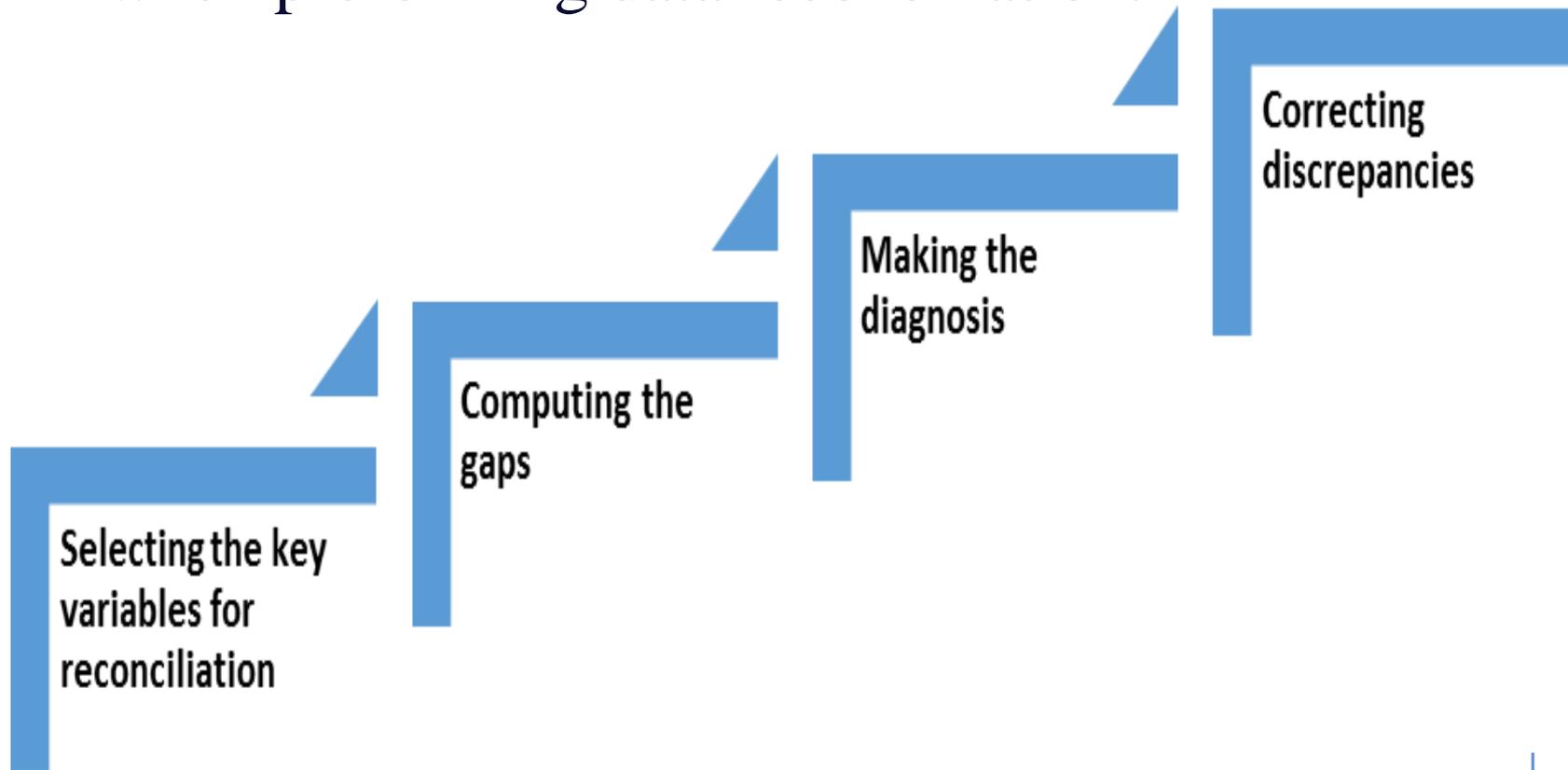


### **3. Common sources of discrepancies between Census and Surveys data (Cont'd.)**

- When sample surveys are conducted, discrepancies with census data could be ascribed to the disappearance, division, or merger of holdings over time due to endogenous or exogenous events.
- Phenomena occurring in the population may also impair sample quality.
- These changes adversely affect the quality of panel of holdings because they directly influence sample size and the weight of the statistical units.

## 4. Correcting discrepancies - Data reconciliation path

- Four main operational steps may be considered when performing data reconciliation:



## 4. Data Reconciliation Path (Cont'd.)

### Step 1: Selecting the key variables for reconciliation

- Data reconciliation may be a time consuming task.
- Rigorous data reconciliation may not be possible for all the survey variables. Therefore, it is important to identify a number of key variables for the reconciliation process.

### Step 2: Computing the gaps

- The second step is to calculate the gap between surveys and census data regarding the key variables selected.
- Both absolute and relative growth needs to be made available for the next step.
- Comparisons of ratios (e.g. proportion of maize planted area) may also be helpful given that ratios are not likely to change in the short term.

## 4. Data Reconciliation Path (Cont'd.)



### Step 3: Making the diagnosis

- For each key variable, analyse if the gap is normal or if there is a discrepancy.
- Some gaps may be linked to the normal evolution of variables from the year of the survey to the one of the census.
- In some cases, previous conjectural factors occurred in the country may explain the differences.
- In this step the opinions of experts with deepen knowledge of the agricultural economy of the country may be helpful.
- Secondary data may also help to understand some gaps.
- When discrepancies are identified, their sources should be explored in order to assess the causes of the gaps.

## 4. Data Reconciliation Path (Cont'd.)



### Step 4: Correction

- After the diagnosis, it follows the correction of discrepancies.
- For each source of gaps identified, suitable methods to correct discrepancies should be considered.
- The Global Strategy (2016) discussed some methods used in the literature for data reconciliation.

## 5. Methods used for data reconciliation



- 1. Source of gap:** Misclassification of units in strata
  - **Method:** Post-stratification.
  - **Description:** Post-stratification helps to adjust the survey sample data to make it more consistent with the population's structural parameters based on the census data.
  
- 2. Source of gap:** Some holdings were overrepresented in the sample
  - **Method:** Ad hoc trimming method.
  - **Description:** The method establishes an upper cut-off point for large weights, reducing weights larger than the cut-off weight to the cut-off value and then redistributing the weight in excess of the cut-off to the non-trimmed cases. This ensures that the weights before and after trimming add up to the same totals.



## 5. Methods used for data reconciliation (Cont'd.)

3. **Source of gap:** Estimation errors due to ageing of sampling weights: sampling weights correction.
  - **Methods:** (1) Best Linear Unbiased Prediction (BLUP) method, (2) Robust BLUP method and (3) Difference estimator method.
  - **Description:** These model-based methods re-estimate the survey data through new sampling weights calculated using a model. The model estimates the values of a target variable for unobserved units using the census data and auxiliary variables.
  
4. **Source of gap:** Estimation errors due to ageing of sampling weights: sampling weights correction.
  - **Method:** Growth rate method.
  - **Description:** Growth rate is estimated using many years of census data and used to adjust the sampling weights. The new estimates are inflated or deflated according to the growth rate. 13

## 5. Methods used for data reconciliation (Cont'd)

### 5. **Source of gap:** Errors in weights.

- **Methods:** (1) Cross-entropy estimation method, (2) Generalized Regression (GREG) method and (3) Spline method (Robust GREG).
- **Description:** These methods propose adjustments of the sampling weights using the new census data or other new and accurate estimation. It makes the sample resemble the new population, but at the same time maintains the adjusted weights as similar to the original weights as possible.

### 6. **Source of gap:** Misclassification of members and non-members of the population.

- **Method:** Capture-recapture approach.
- **Description:** This method is applied when the census and survey are performed in the same reference period. It estimates the probability that a surveyed holding is captured by the census and it is used to adjust the original sampling weights.

## 6- Country example

### Data reconciliation by Statistics Canada

#### 1. Census data validation process

- Objectives: to guarantee the quality and consistency of the census of agriculture data and to make recommendations for their publication.
- It is a complex process in which human judgement is vital
- The process starts at the more general/macro level (i.e. provincial overviews), flows towards the more specific/micro level (i.e. individual census questionnaires) and then ends with a final macro-level reviews.
- Respondent data may be compared to previous surveys' data of the same individuals in cases where they had been sampled.

## 6- Country example (Cont'd.)

### Data reconciliation by Statistics Canada

#### 2. Intercensal revision of surveys data

- Revisions usually completed 1-2 years after the census data are released.
- Survey estimates are revised to match as closely as possible the census numbers, being adjusted for seasonal variation when appropriate.
- The revisions made on commodity data can be summarized by being either a wedge adjustment or a logarithmic adjustment.
- Only the trend is adjusted, not the magnitude of the change year to year.
- Comparison of variables (e.g. area) is first done between surveys and the census to determine the extent of frame change and potential intercensal adjustments.
- Ratios are used in different ways for each commodity to support the analysis: **(a)** the ratio of published numbers to census numbers; **(b)** the ratio of census numbers compared to survey-level estimates; **(c)** the ratio of average yield (from the survey) on the total area (from the census) to adjust production; **(d)** the census livestock inventory data adjusted for seasonal variation (for cattle and sheep), etc.

## 6- Country example (Cont'd.)

### Data reconciliation by Statistics Canada

#### **Certification:**

- Revised survey estimates are verified by other members of the team. Provincial experts are also consulted to have their viewpoints on the possible extent of revision.

#### **Communication plan:**

- A communication plan is put in place to inform all key users that new intercensal revisions have been made available. Typically, users know that estimates are revised every five years.



**Thank you**