



This project is funded  
by the European Union



Schweizerische Eidgenossenschaft  
Confédération suisse  
Confederazione Svizzera  
Confederaziun svizra

Swiss Cooperation Office Albania  
Zyra e Bashkëpunimit Zviceran në Shqipëri



# Quality dimensions of the 2011 Albanian Census

Guido Pieraccini

# Dimension of Quality

- The UNECE *Recommendations for the 2010 Censuses of Population and Housing*, consider quality assurance as a multi-dimensional concept, including different dimensions.
- In that approach, *Data Accuracy* is only one – albeit important – dimension of the overall data quality.
- The quality publication is divided in two main parts: a first one - more qualitative - that provides a detailed overview, together with some elements of evaluation, about the main measures implemented to ensure the quality of the Census process and a second one - more quantitative - focused on data Accuracy.

# Quality assurance measures in 2011 PHC

## Relevance

Meet users' needs

---

## Completeness

Serve user needs as completely as possible

---

- Compliance with the information requirements of the Census law
- Consultations with data users and Census experts on questionnaire content and dissemination program
- Production high-quality field maps that increase the likelihood of full coverage
- Application of effective system for monitoring the enumeration progress

## Timeliness

Figures up-to-date and published on time

---

- Training of a pool of reserve field staff to avoid delays in enumeration due to drop-out
- Effective transport and storage procedures for completed Census materials
- Development of rapid data processing procedures to compile timely provisional results

# Quality assurance measures in 2011 PHC

## Coherence

Common definitions, classifications and methodological standards

---

## Comparability

Enable comparisons across space and time

---

- Compliance with international recommendations on definitions, 'core topics' and classifications
- Overall comparability with the 2001 Census and with targeted surveys (LSMS, LFS)
- Inclusion of questions used to implement the 2012 Agricultural Census
- Adopting of a new sample frame based on the Census data

## Accessibility

Figures easily accessible in the form users need

---

- Implementation of a communication strategy to inform people about the Census
- Implementation of a Census website
- A sample of anonymized Census micro data is available on the Census website

# Quality assurance measures in 2011 PHC

## Accuracy

Degree to which the data correctly estimate the characteristics that the survey is measuring

---

- Consultations with Census experts to assure the correct conceptualization of the data to be collected
- Test of questionnaires, forms, and procedures in a field test and in a Pilot Census
- A team of national and international experts was set up to support the coordination and the monitoring of the field operations, both for Census and post-enumeration survey
- Assessment of the quality of the editing and imputation procedure
- Conduction of a post-enumeration survey to estimate the Census coverage

# Census fieldwork observation

- Logistic preparation and training received by supervisors and controllers was considered very favorable.
- Field staff dropping out was particularly prevalent during the training but also noticeable during the enumeration. In most cases, these drop-outs were replaced by persons from the reserve list but some were newly recruited staff.
- The large majority of the enumerators was considered good and committed to their work. A small proportion performed in a substandard way and only a few were totally unacceptable.
- The completed forms were of a good standard. Only a small minority was considered of unacceptable quality.
- No structural problems concerning the refusals was observed.
- Common problems of lower response rates occurred in some areas of Tirana. The opportunity to continue the enumeration in the extension week apparently limited this situations.

# Editing principles

- Not valid values can be replaced through deterministic or probabilistic imputation.
- A deterministic imputation is generated by constrains, which, if violated, lead to an error we are able to correct with a high level of confidence. Deterministic corrections can be performed with a if-then procedure.

**6** What is your country of citizenship?

Multi - tick question

1  Albanian

2  Other Specify

A	L	B	A	N	I	A				
---	---	---	---	---	---	---	--	--	--	--

3  NONE

## Slide 7

---

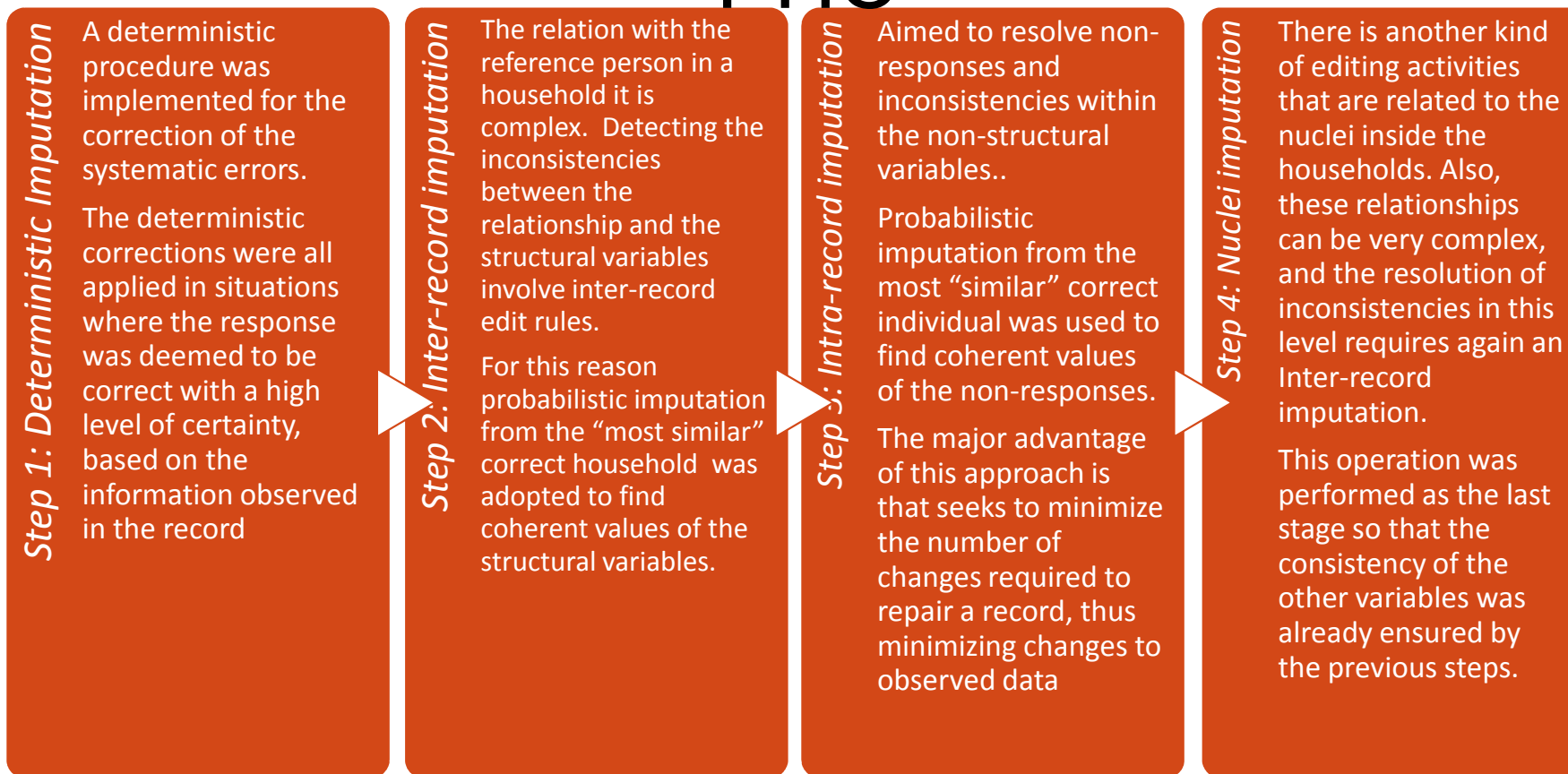
- G6** As any statistical operation, Census is subjected to errors, resulting in a certain amount of data not valid.  
Guido, 20/05/2014
- G7** Non-responses include missing, multi-ticks and out of range values.  
It is also common for some in-range values to be considered invalid because they are inconsistent with other values on the questionnaire. Inconsistencies are detected by validating the data against a set of pre-defined edit rules.  
Guido, 20/05/2014
- G8** Non-response and inconsistency can lead to bias and contradictions in the analysis, so it is important to correct them.  
This is desirable because users do not have all the information needed to estimate Non-response and inconsistencies and because it prevents the adoption of different methods of editing, which could result in contradictory results  
Guido, 20/05/2014



# Editing principles

- A probabilistic imputation is generated by a rule that, if violated, lead to an error that we are not sure how to correct.
- An example of probabilistic edit is “civil status is married imply that age is more than 15”. When this rule is violated, we know that the situation is wrong, but we are not able to identify clearly where the error is: could be that the age is wrong or could be that the civil status is wrong.
- Donor-based methods built a deck of donors from the set of records that shows no violations of the edit rules (clean records) and looks for the minimum distance between the erroneous record and records in the donor’s deck. Once identified the closest clean record, its values are used to replace the item non-responses of the erroneous records.

# Editing and imputation strategy for the 2011 PHC



# Editing and imputation assessment

	Individual dataset	Household dataset	Dwelling dataset
Number of Records	2,800,138	722,262	1,021,332
Number of Variables	66	30	9
<b>Number of Total values</b>	184,809,108	21,667,860	9,191,988
<b>Number of Valid values</b>	180,558,499	21,376,712	9,003,207
<b>Number of Imputed values</b>	4,250,609	291,148	188,781
<b>Number of Additions</b>	2,180,747	236,641	117,126
<b>Number of Eliminations</b>	1,201,259	49,588	7,272
<b>Number of Modification</b>	868,603	4,919	64,383
<b>Imputation rate (I)</b>	<b>2.3</b>	<b>1.3</b>	<b>2.1</b>
Additions rate (Ia)	1.2	1.1	1.3
Elimination rate (Ie)	0.6	0.2	0.1
Modification rate (Im)	0.5	0.0	0.7
<b>Non-Imputation rate</b>	<b>97.7</b>	<b>98.7</b>	<b>98.0</b>
<b>% of records with I greater than 2%</b>	30.9	16.0	14.2
<b>% of records with I greater than 5%</b>	11.5	5.2	14.2

# Editing and imputation assessment

	Imputation Rates				Dissimilarity Indexes	
	l	la	le	lm	Type	Value
Citizenship: Albanian	10.98	10.96	0.01	0.00	IM	0.110
Had live-born children	9.81	5.47	4.19	0.15	IM	0.013
Family nucleus	9.49	3.80	0.00	5.69	IM	0.058
Completed years of education	9.06	1.05	2.24	5.76	KS	0.012
Highest completed level of education	7.55	1.13	2.05	4.37	IM	0.022
Searched for work during the month of September	5.76	1.70	3.56	0.50	IM	0.019
Place of residence in 2001	5.72	4.02	1.38	0.32	IM	0.026
Had a job last week of September	5.31	1.29	4.02	0.00	IM	0.027
Willingness to start job in two weeks	4.92	1.33	3.59	0.01	IM	0.023
Place of work: Type	4.57	3.78	0.29	0.50	IM	0.035
Place of residence was changed	4.13	2.15	0.14	1.84	IM	0.020
Have lived abroad	4.10	1.94	1.97	0.19	IM	0.020
Worked on last week of September	4.08	1.15	2.15	0.78	IM	0.019
Employment Status	4.07	2.47	0.00	1.60	IM	0.015
Currently attending formal education	3.79	2.99	0.45	0.34	IM	0.028
Place of work	3.69	3.23	0.33	0.12	IM	0.029
Main reason for not searching for work	3.41	0.21	3.18	0.02	IM	0.030
Place of stay on CRM	3.08	1.25	0.00	1.83	IM	0.019

# The Post Enumeration Survey

- Even though the word 'Census' implies a 100 per cent count of the people and housing, it is rarely possible to achieve this.
- The Conference of European Statisticians (CES) agreed in their preparations for the 2010 and 2011 Censuses in Europe that the Censuses should be evaluated to check on coverage and the quality of the information provided. <sup>G9</sup>
- For the first time in the history of Censuses in Albania, INSTAT conducted a post enumeration survey to measure the effectiveness and reliability of the 2011 Census
- The post enumeration survey had good preparation and was executed effectively and efficiently, though late planning at various stages and the use of same staff of the Census were noted.

## Slide 12

---

**G9** Further, the European legislation on the censuses requires Member States to report on the quality of the census results.

Guido, 20/05/2014

# Estimation of the population undercount

*Percentage undercount of population*

Albania	3.0
In urban areas	4.0
In rural areas	1.8

- The PES data cannot give indications about refusals, but estimation of them can be derived from the Census data.
- The total amount of dwellings that refused the Census is 6,765. On this base, the number of persons who refused to participate to the Census is estimated in 21,839 (0.8%).
- Under the reasonable assumption that who refused the Census also refused the PES, the total undercounts of the population became 3.8%. If we don't want to consider this assumption we still can say that the total undercount is somewhere between 3.0% and 3.8%
- The measure of the undercount of population, when taken with the non-response to the Census, is acceptable and within limits achieved by developed countries.

# Evaluation of the Census quality based on Census/PES matched records

*Percentages of differences in the Census/PES matched records*

	Different %	Equal %
Sex	1,3	98,7
Civil Status	3,8	96,2
Age	6,9	93,1

- Both the Census and PES operations are affected by errors. Not all the difference between Census and PES data should be automatically considered as a mistake.
- Assumption that mistakes in PES and Census are equally distributed

*Indicators of Total error and Total quality for Sex, Civil status and Age*

	TE	TQ
Sex	0,7	99,3
Civil Status	1,9	98,1
Age	3,5	96,5



# Conclusion

- Various field problems were identified during the enumeration but, in relation to the number of staff involved and the overall scale of the operation, had not major implications on the Census results.
- A full evaluation of the editing and imputation procedure has been undertaken and published. The analysis shows a general low impact of the edit and imputation procedure on the data related to the different items. In a few cases a more relevant impact was observed, but at the same time, the original distributions have been always preserved.
- The PES had good preparation, was executed effectively and efficiently and the processing supervised by international experts. The measure of the undercount of population is acceptable and, when taken with the non-response to the Census, is within limits achieved by developed countries.