



INSTITUTO NACIONAL DE ESTATÍSTICA  
STATISTICS PORTUGAL

**CENSOS**  **2021**

# When is administrative data good enough to replace statistical information? A quality indicator based on census comparison

Meeting of the Group of Experts on Population and Housing Censuses

28-30 September 2016, Geneva

Anabela Delgado<sup>1</sup>, Sandra Lagarto<sup>2</sup>, Paula Paulino<sup>3</sup>, João Capelo<sup>4</sup>

<sup>1,2,3,4</sup> Statistics Portugal

Email: <sup>1</sup>[anabela.delgado@ine.pt](mailto:anabela.delgado@ine.pt) <sup>2</sup>[sandra.lagarto@ine.pt](mailto:sandra.lagarto@ine.pt)

<sup>3</sup>[paula.paulino@ine.pt](mailto:paula.paulino@ine.pt) <sup>4</sup>[joao.capelo@ine.pt](mailto:joao.capelo@ine.pt)

# Outline of presentation

## BACKGROUND

- 2021 Portuguese Census Programme
- Legal basis

## ADMINISTRATIVE DATA

- Sources selection
- Variables selection

## METHODOLOGY

- Equality rate 2011 Census – Administrative data
- 2011 Census Post Enumeration Survey (PES)
- Equality rate 2011 Census – 2011 Labour Force Survey (LFS)

## RESULTS AND DISCUSSION

- Comparison 2011 Census microdata – Administrative registers
- 2011 Census PES (ICG)
- Comparison 2011 Census microdata – 2011 LFS microdata

## CONCLUSIONS



want



need



have



done



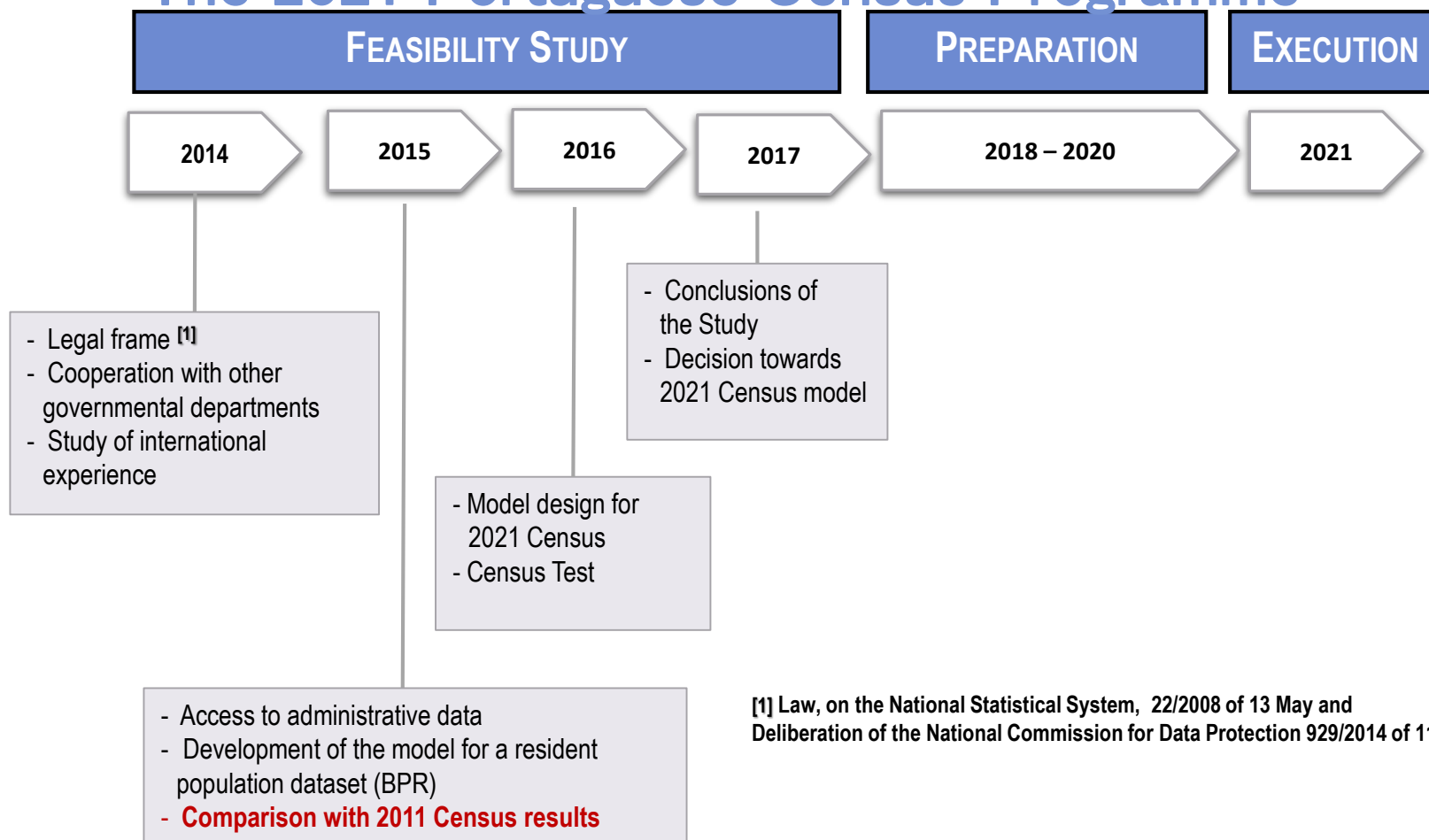
got



next

# Go to an administrative census model

## The 2021 Portuguese Census Programme



[1] Law, on the National Statistical System, 22/2008 of 13 May and Deliberation of the National Commission for Data Protection 929/2014 of 11 June



want



need



have



done



got



next

# Administrative sources needs

## CENSUS QUALITY STANDARDS ADMINISTRATIVE DATA WHICH ALLOW:

- To apply record linkage/matching techniques to **estimate resident population**
- To **cover all resident population topics**: core and other national optional topics (non-core)



want



need



have



done



got



next

# Available administrative sources

SOURCE NAME	YEAR	NO. RECORDS	DESCRIPTION	DATASET NAME
Institute of Registration and Notary	2010	11 565 714	PORTUGUESE CIVIL REGISTER	BDIC
Immigration and Borders Service	2011	434 708	FOREIGNER REGISTER	SEF
Social Security Institution	2011	7 209 027	SOCIAL SECURITY REGISTER	ISS
Strategy and Planning Office	2011	2 736 659	PRIVATE EMPLOYMENT REGISTER	QP
Institute of Employment and Training and Regional Directorate of Statistics of Madeira	2011	702 215	UNEMPLOYMENT REGISTER	IEFP
General Directorate of Education and Science Statistics and Regional Secretariat for Education and Human Resources of the Autonomous Region of Madeira	2011	1 965 842	SCHOOL AND UNIVERSITY REGISTER	EDUC
General Retirement Fund	2010	1 103 980	PUBLIC ADMINISTRATION RETIREMENT FUND REGISTER	CGA



# Available administrative variables

DATASET	AVAILABLE INFORMATION ON POPULATION TOPICS
PORTUGUESE CIVIL REGISTER	Place of residence (municipality), sex, date of birth, legal marital status, country of birth, country of citizenship
FOREIGNER REGISTER	Country of birth, country of citizenship, current activity status, occupation
SOCIAL SECURITY	Current activity status, place of work, status in employment
PRIVATE EMPLOYMENT	Place of work (municipality), occupation, industry (establishment), status in employment, number of persons working in the enterprise, hours usually worked, educational qualifications
UNEMPLOYMENT	Current activity status
PUBLIC RETIREMENT FUND	Current activity status
EDUCATION	School attendance



want



need



have



done



got



next

# Methodology

<b>PROBLEM</b>	Can we use administrative data to replace data collected by census? If we could, can we rely on them?
<b>HYPOTHESIS</b>	If administrative data are equal to census collected data, we can rely on them for census statistics.
<b>GOAL</b>	To compare, for each person, the exact value of the target census variables, available on administrative datasets.
<b>How?</b>	Estimate an equality rate, comparing the 2011 Census -administrative datasets variables.
<b>ASSUMPTIONS</b>	Data is up to date, there´s no content errors.
<b>REQUIREMENTS</b>	Match Census records to administrative registers.
<b>RESULTS VALIDATION</b>	Use a quality indicator of 2011 Census PES and the comparison results of 2011 Census – 2011 LFS variables.



want



need



have



done



got



next

# Comparison results: geographical and demographic variables

VARIABLE	2011 CENSUS POPULATION TO BE COMPARED	NUMBER OF ADMINISTRATIVE RECORDS TO BE COMPARED TO 2011 CENSUS, BY SOURCE		NUMBER OF PAIRS COMPARED	EQUALITY RATE ON COMPARED PAIRS (%)	ICG <sup>[2]</sup> (%)
PLACE OF RESIDENCE (MUNICIPALITY)	10 562 178	11 565 714	BDIC	9 308 384	94,6	97,7
SEX	10 562 178	11 565 714	BDIC	9 308 384	99,9	99,0
DATE OF BIRTH	10 562 178	11 565 714	BDIC	9 308 384	92,6	95,7
LEGAL MARITAL STATUS	10 562 178	11 565 714	BDIC	9 308 384	95,3	97,4
COUNTRY OF BIRTH	10 562 178	11 565 714	BDIC	9 308 384	94,7	84,0
		434 708	SEF	107 136	91,3	84,0
COUNTRY OF CITIZENSHIP	10 562 178	11 565 714	BDIC	9 308 384	99,4	97,8
		434 708	SEF	107 136	90,3	97,8

**[2] ICG measures content errors; it represents the percentage of statistical units – in this case resident population –, with the same classification both in 2011 Census and 2011 Census PES, of all common units to the two statistical operations**

Ref.: INE – Instituto Nacional de Estatística (2013), *Inquérito de Qualidade dos Censos 2011 – Metodologia e resultados*, Instituto Nacional de Estatística, Lisboa

# Comparison results: economical and educational variables

VARIABLE	2011 CENSUS POPULATION TO BE COMPARED	NUMBER OF ADMINISTRATIVE RECORDS TO BE COMPARED TO 2011 CENSUS, BY SOURCE		NUMBER OF PAIRS COMPARED	EQUALITY RATE ON COMPARED PAIRS (%)	ICG (%)
CURRENT ACTIVITY STATUS	8 989 849	7 066 838	ISS	4 910 073	81,2	
		379 965	SEF	107 136	27,1	
		1 103 980	CGA	716 264	92,1	
		702 215	IEFP	454 479	42,1	
PLACE OF WORK	4 361 187	4 107 425	ISS	2 788 758	56,6	77,6
		2 736 659	QP	2 045 476	81,6	77,6
OCCUPATION	4 361 187	2 736 659	QP	2 045 476	61,9	
		124 721	SEF	171 370	52,9	
INDUSTRY	4 361 187	2 736 659	QP	2 045 476	74,1	
STATUS IN EMPLOYMENT	4 361 187	2 736 659	QP	2 045 476	93,0	82,2
		4 107 425	ISS	2 788 758	85,5	82,2
NUMBER OF PERSONS WORKING IN THE ENTERPRISE	4 361 187	2 736 659	QP	2 045 476	54,4	51,6
HOURS USUALLY WORKED	4 361 187	2 736 659	QP	2 045 476	56,8	
EDUCATIONAL ATTAINMENT	10 445 093	2 736 659	QP	2 210 930	59,5	
SCHOOL ATTENDANCE	10 445 093	1 965 842	EDUC	1 359 916	82,2	69,8

# Comparison results: 2011 Census – Administrative registers vs. 2011 Census – 2011 LFS

NUMBER OF PAIRS COMPARED (2011 CENSUS – 2011 LFS<sup>[3]</sup>)

17 732

VARIABLE	EQUALITY 2011 CENSUS – 2011 LFS (%)	% EQUALITY CENSUS – ADMINISTRATIVE RECORDS, BY SELECTED ADMINISTRATIVE DATA SOURCE	
LABOUR FORCE STATUS	84,3	81,2	ISS
OCCUPATION	67,8	61,9	QP
INDUSTRY	77,6	74,1	QP
STATUS IN EMPLOYMENT	86,5	93,0	QP
NUMBER OF PERSONS WORKING IN THE ENTERPRISE	60,6	54,4	QP
HOURS USUALLY WORKED	72,6	56,8	QP
EDUCATIONAL ATTAINMENT	80,2	59,5	QP
SCHOOL ATTENDANCE	86,5	87,4	EDUC

[3] The Portuguese LFS, which is conducted nationwide, is a sample survey providing quarterly results (recently monthly). Back in 2011, it collected labour market information for approximately 40 000 individuals.

# Main conclusions

## Results show huge consistency between administrative data and 2011 census microdata

- Equality rates on the 8 demographic variables compared are very high (90 per cent of the compared pairs of records' information are exactly the same)
- For some of the 10 labour force related and educational variables compared, equality rates show more than 80 per cent of equality
- QP is the most consistent source of administrative information with, globally, the highest equality rates

## Comparability indicators show inequalities only based in unequal values; so, administrative data can be use to add or replace information collect by census

- Time lags between datasets and some conceptual issues could explain differences on comparison results
- Data sources holders are being contacted for new data flows and we believe that some of the issues that cause inequalities can be solved with more recent incomes

## The reliability of using administrative data for statistical purposes was confirmed by using additional quality information criteria from PES and 2011 Census vs. 2011 LFS comparison results

Even though there are no doubts about usability potential of the Portuguese administrative data for census purposes, it is obvious that some variables are not fully or even partially covered by the currently available data (e.g. some core topics related with household or education)







want



need



have



done



got



next

# Next steps

## Work in progress/ future work

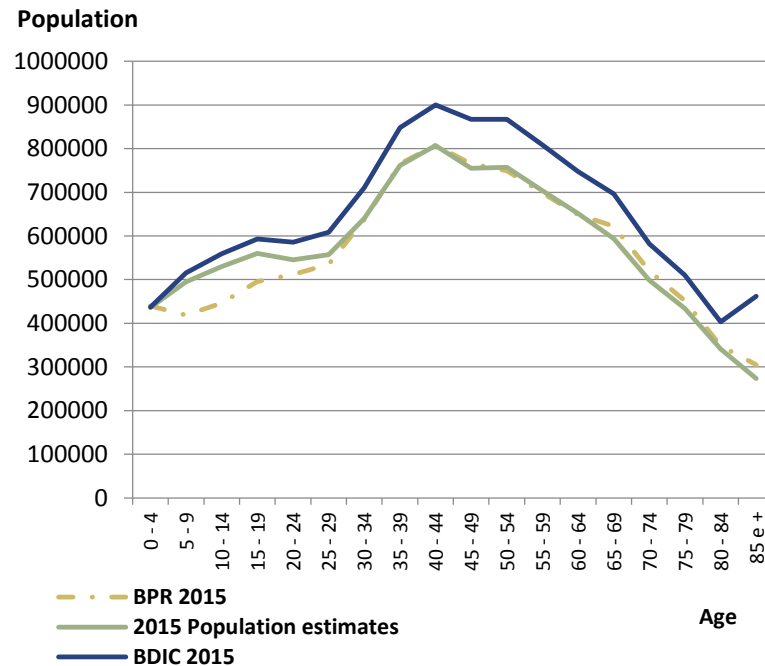
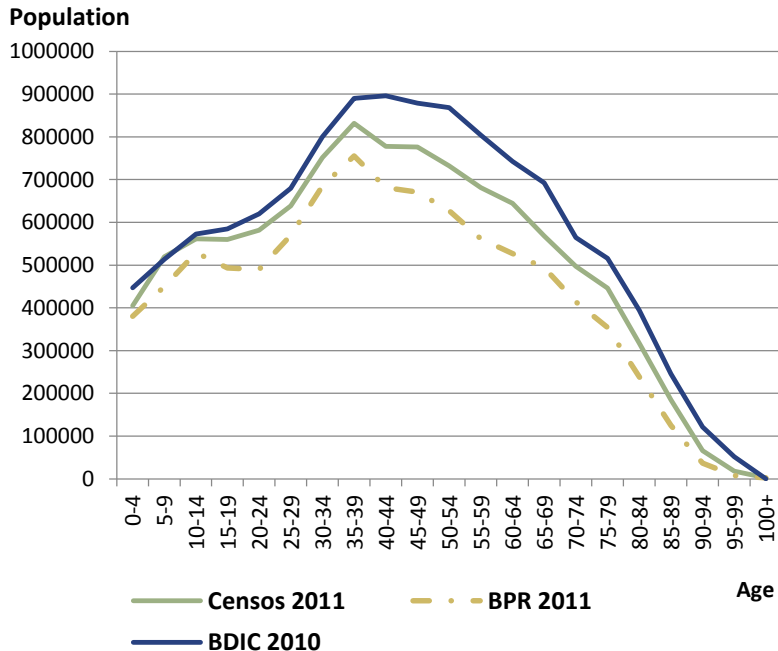
- Cross comparison and hierarchical rules between sources of administrative information are being studied
- Repeat the comparison exercise with 2015 Census Test results and recent incoming administrative datasets

# Freshest results in a glance

## Estimates of the resident population dataset (BPR) based on signs of life methodology

2011 Census: 10,5 million people  
 2011 Estimates of the resident population dataset (BPR): 9,0 million  
 Civil register dataset (BDIC) : 11,5 million  
**Census – BPR  $\equiv$  1,5 million (underestimation rate: 14%)**

2015 Population estimates: 10,3 million people  
 2015 BPR : 10,1 million  
**Estimates – BPR  $\equiv$  200 000 (underestimation rate: 2%)**





# Thank you!



[joao.capelo@ine.pt](mailto:joao.capelo@ine.pt)  
[sandra.lagarto@ine.pt](mailto:sandra.lagarto@ine.pt)  
[anabela.delgado@ine.pt](mailto:anabela.delgado@ine.pt)  
[paula.paulino@ine.pt](mailto:paula.paulino@ine.pt)

