

Evaluation of census data with Health and Demographic Surveillance Systems: a record linkage study in rural Senegal

Bruno Masquelier^{1,2}, Cheikh Tidiane Ndiaye³, Gilles Pison²,
Ndeye Binta Dieme³, Samba Ndiaye³, Ousmane Ndiaye⁴, Lucie
Lecomte⁴, Valérie Delaunay⁴

¹Centre for Demographic Research (UCL, Belgium)

²Institut National d'Etudes Démographiques (INED, France)

³Agence Nationale de la Statistique et de la Démographie (ANSD, Senegal)

⁴Institut de recherche pour le développement (IRD, France)

INDEPTH scientific conference, Addis Ababa, 11-13 November
2015

Introduction

Data and methods

Comparisons at the aggregate level

- Age structures

- Mortality levels and trends

- Age-specific fertility rates

Linkage rates and quality of matches

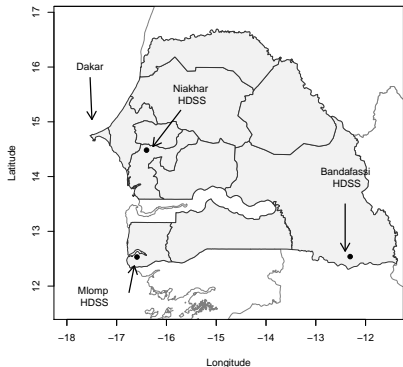
Conclusion

Introduction

- ▶ In the absence of complete CRVS systems, censuses will remain a fundamental data source on population dynamics in low-income countries.
- ▶ Health and Demographic Surveillance Systems (HDSS) can help in assessing the quality of cross-sectional demographic data collected in censuses.
- ▶ We extracted from the 2002 and 2013 censuses in Senegal the data covering three HDSS and matched both datasets at the individual level.
- ▶ Here we present preliminary results from aggregate and individual comparisons.

Data and methods

3 HDSS in Senegal



- ▶ Bandafassi, near the borders with Mali and Guinea, established in 1970, now covering 42 villages and 13k inhabitants.
- ▶ Mlomp, in Casamance, near the border with Guinea-Bissau, established in 1985, now covering 8k inhabitants in 11 villages.
- ▶ Niakhar, 135km southeast of Dakar, established in 1962, now covering 30 villages with 44k inhabitants.

2 censuses: 2002 and 2013



- ▶ Census 2002: paper-based, conducted between 8 & 22 December
- ▶ Census 2013: handheld computers (PDA), conducted between 19 Nov & 9 Dec

Most of questions recommended by the United Nations (2008): total number of children ever born and surviving, orphanhood, births and deaths in the last 12 months + migrations and birth/death registration.

Importance of the definition of residence: in most cases, anyone who has stayed or intent to stay in the compound for *6 months* or more.

Methods

1. Comparison of names of villages/hamlets and harmonisation between HDSS and censuses.
2. Identification in HDSS databases of all residents at the time of the two censuses.
 - ▶ Based on data collected on seasonal migrations (in Mlomp and Niakhar only), exclusion of HDSS residents which would not be considered as residents in the censuses.
3. Extraction of census data pertaining to the HDSS areas.
4. Analysis of age structures, trends in fertility and mortality based on indirect techniques.
5. Record linkage between the two data sources.

Record linkage strategy

Three steps:

1. First automatic search based on first/last names.
2. Manual linkages based on kinship charts.
3. Field verification with local informants and HDSS interviewers.

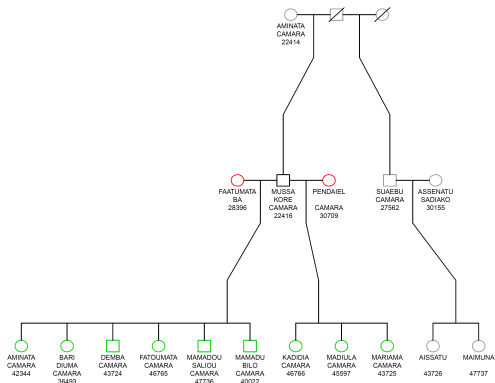
No use of data on ages, number of children ever born, etc. because the quality of these reports are analyzed.

[1] More details on the automatic search

- ▶ Spelling of the most common first and last names is harmonized between censuses and HDSS.
- ▶ Each household from the census is compared to all households in the same village (and neighboring villages) in the HDSS.
- ▶ First and last names are compared based on the Jaro-Winkler distance.
 - ▶ Ex: BIRAM vs BIRAME= 0.94, ABLAYE vs ABDOULAYE= 0.89, BIRAM vs ABDOULAYE= 0.37.
- ▶ Two names are considered identical if the score is ≥ 0.9 (a fairly strict criterion).
- ▶ Two households are provisionally matched if at least half of their members have a similar first+last name (and their size does not differ too much).

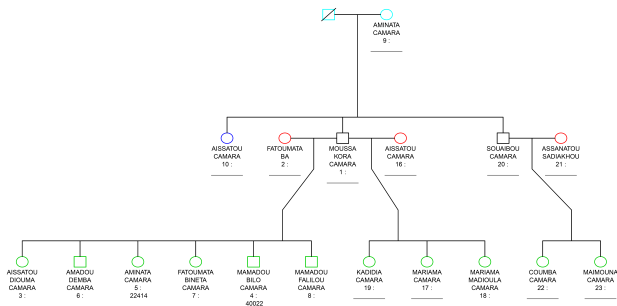
[2] More details on the manual linkages

Kinship charts are drawn based on relationship to head (and order of listing) in the censuses, and genealogical data in the HDSS. Linkages based on automatic search are here as a guide only.



HDSS kinship chart
Note: names have been changed.

Census kinship chart (Bandafassi 2013)

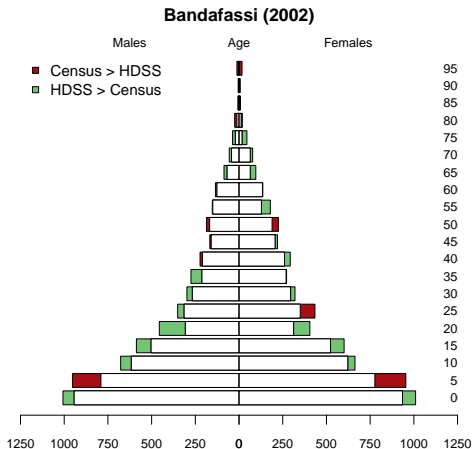


A single potential match found automatically can help to match many other household members (here we clearly see the same structure between HDSS and census).

Manual linkages are made in the office and kinship charts are updated for the field verification.

Results

Comparison of age structures

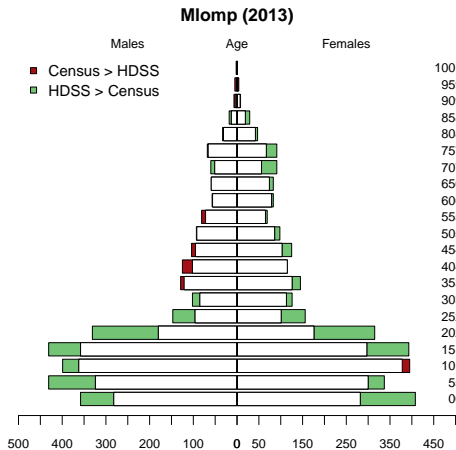


Differences in population size and age structures might come from

- ▶ coverage and content errors (age misreporting).
- ▶ various definitions of residence.

In Bandafassi, there is a better agreement between HDSS and census data than in Mlomp and Niakhar.

Comparisons at the aggregate level Age structures



- ▶ Even after excluding seasonal migrants, the HDSS tend to include more residents than the censuses.
- ▶ Differences are particularly large among young adults (15-24), indicating that some seasonal migrants could not be identified in the HDSS databases.

Relative differences in enumerated populations

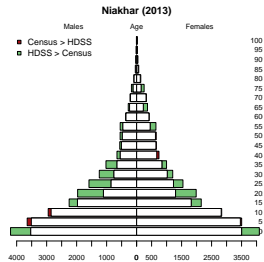
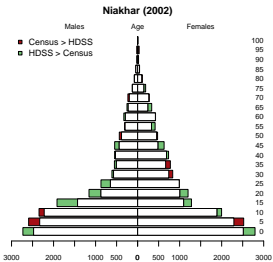
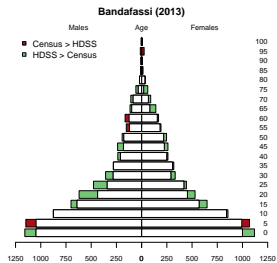
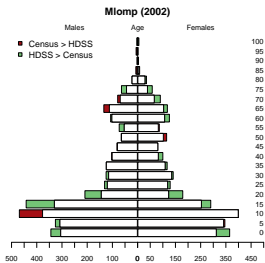
	Bandafassi		Mlomp		Niakhar	
	Enumerated pop. ¹	Relative difference ²	Enumerated pop. ¹	Relative difference ²	Enumerated pop. ¹	Relative difference ²
2002						
0-15	5016	+ 1.9 %	2133	- 0.9 %	14332	- 0.2 %
15-60	4966	- 9.3 %	2219	- 13.5 %	12393	- 11.4 %
60 +	608	- 15.1 %	733	- 10.3 %	2313	- 4.6 %
2013						
0-15	5975	- 1.2 %	1939	- 16.2 %	19946	- 4.8 %
15-60	5616	- 12.6 %	2414	- 20.4 %	16025	- 23.2 %
60+	813	- 15.4 %	625	- 13.5 %	2396	- 14.6 %

1: According to the census - 2: Estimated as $(n_{census} - n_{hdss})/n_{hdss}$

Differences are larger for adults aged 15-60 and they increase overall from 2002 to 2013.

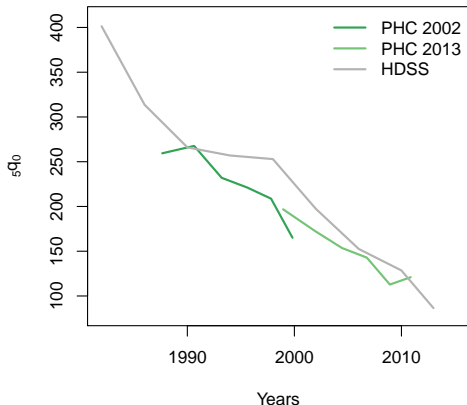
Comparisons at the aggregate level

Age structures



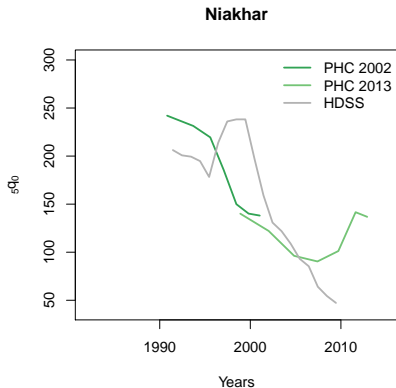
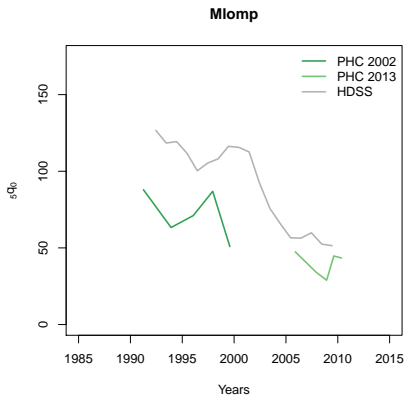
Mixed results when comparing mortality indicators

Bandafassi



- ▶ Trends in child mortality obtained from questions on children ever born and surviving in the census.
- ▶ In Bandafassi, good agreement with trends obtained from HDSS, although $5q_0$ is approximately 10% too low in the censuses.

Mixed results when comparing mortality indicators



In the other HDSS, indirect estimates are unreliable (too low in Mlomp, recent trends implausible in Niakhar).

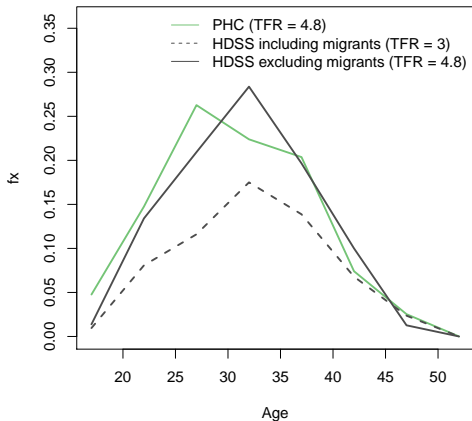
High quality of data on recent household deaths

Most mortality indicators obtained from recent household deaths are in agreement with HDSS, except for infant mortality in 2002.

		Bandafassi		Mlomp		Niakhar	
		HDSS	Census	HDSS	Census	HDSS	Census
2002	1q0	0.11	0.04	0.05	>0.01	0.06	0.02
	4q1	0.09	0.11	0.04	0.05	0.08	0.08
	45q15	0.25	0.21	0.54	0.59	0.26	0.24
	20q60	0.68	0.71	0.64	0.43	0.52	0.49
	e5	61.9	62.9	45.7	43.5	64	64.5
2013	1q0	0.06	0.08			0.02	0.02
	4q1	0.04	0.04			0.02	0.03
	45q15	0.29	0.38			0.13	0.19
	20q60	0.59	0.48			0.52	0.52
	e5	62.4	61.3			68.7	66.6

Importance of identifying seasonal migrants

Mlomp 2002



- ▶ Necessary to isolate seasonal migrants when comparing fertility rates, otherwise the comparison with the census is biased.
- ▶ Women registered as residents in the census are more fertile than all women in the HDSS database (here in Mlomp).

Linkage rates by site (%): population enumerated in the censuses found in HDSS

	Bandafassi 2013	Mlomp 2002	Mlomp 2013	Niakhar 2002	Niakhar 2013
<i>Automatic search based on names</i>					
Compounds	27.6	33.2	31.7	89.3	86.8
Households	-	24.6	28.2	82.9	77.0
Individuals	11.8	15.0	17.7	43.5	56.1
<i>Manual linkage based on kinship charts</i>					
Compounds	58.0	-	-	95.6	94.2
Households	-	-	-	94.7	90.9
Individuals	45.1	-	-	74.9	77.8
<i>Field verifications</i>					
Compounds	87.7	95.2	86.9		
Households	-	92.2	86.4	<i>in progress</i>	
Individuals	74.8	87.6	83.4		

Matching rates reached in the different steps vary by HDSS, but at the end we can match at least 75% of individuals

Matched and unmatched cases: example of Bandafassi in 2013

	Freq.	Matching rate	Odds ratio	CI's
Sex				
Males	6182	0.75	1	
Females	6226	0.75	1	(0.92-1.08)
Age group				
15-49	4896	0.76	1	
0-4	2045	0.69	0.68	(0.61-0.77)
5-14	3930	0.73	0.85	(0.77-0.93)
50-69	1218	0.84	1.69	(1.43-2.01)
70+	319	0.82	1.38	(1.04-1.86)
Relationship to head				
Head	1369	0.84	1	
Spouse	1768	0.85	1.05	(0.86-1.27)
Children (+in laws)	5996	0.8	0.76	(0.65-0.89)
Grand-children	1260	0.51	0.2	(0.16-0.23)
Other	1763	0.64	0.34	(0.29-0.4)
Unrelated	252	0.2	0.05	(0.03-0.07)

- ▶ Matching rates are higher among adults (especially 50+).
- ▶ No significant association with gender.

An example of individual-level analysis: the quality of reported ages (Bandafassi 2013)

Distribution of age differences between HDSS and census, by age reported at census.

Age (census)	HDSS < Census			=	HDSS > Census		
	10, ...	5-10	2-5		2-5	5-10	10, ...
0-9	0	0.4	2.3	49.3	8.7	2.7	4
10-19	0.7	1.4	2.3	37.5	18.5	7	2.9
20-29	2.4	1.5	3.5	25.7	20.5	12.2	4.6
30-39	1.9	1.9	4.8	32.7	17.6	12.3	6.3
40-49	2.9	2.9	4.8	24.3	17.9	15	8.3
50-59	5.1	5.1	8.4	26.7	13.7	14.2	8
60-69	9.7	7.3	8.7	17	15.2	17.6	6.3
70+	17.2	12.9	8.8	16.3	9.7	7.5	3.8

A tendency to *under-estimate* the ages in the census is apparent in Bandafassi in 2013.

Conclusion

- ▶ Record linkage of HDSS with national censuses is feasible (but time-consuming).
- ▶ Technically quite simple when data on names are available in censuses and genealogical links are known in HDSS.
- ▶ Further analyzes include the quality of reported ages, data on mortality, fertility, etc.
- ▶ Record linkages are also useful to complement HDSS data with census data on disabilities, education, quality of housing, etc.

Thank you!

Questions/comments : valerie.delaunay@ird.fr,
bruno.masquelier@uclouvain.be

References:

United Nations. *Principles and Recommendations for Population and Housing Censuses*. Dept. of International Economic and Social Affairs, United Nations, New York, 2008.