

Using first names to improve the measurement of same-sex couples in the census

Elisabeth Algava et Sébastien Hallépée
Insee, France



A need for reliable statistics

- European Census 2021,
Social and legal recognition of same-sex couples have changed recently

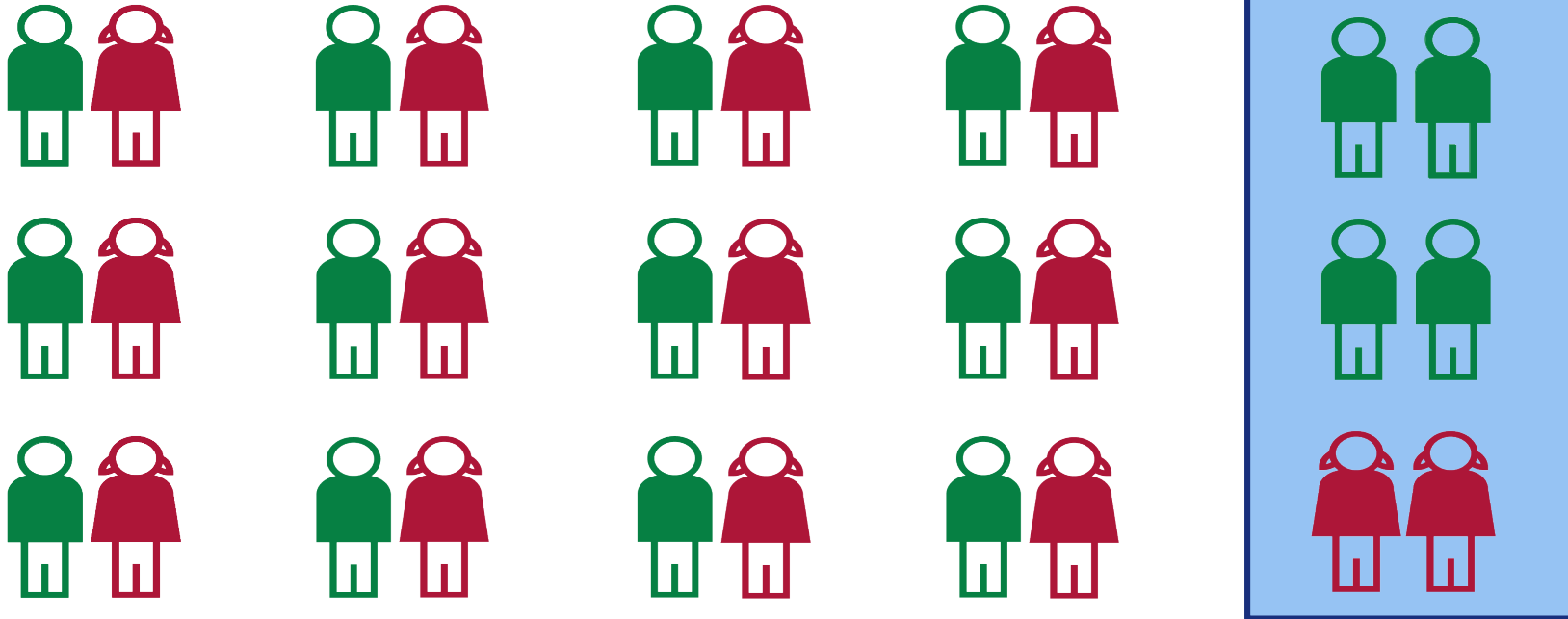
Few data available

- Family and housing survey, 2011 :
 - 0.6% of couples are same-sex couples (SSC)
 - This measure is very sensitive to coding errors on sex.

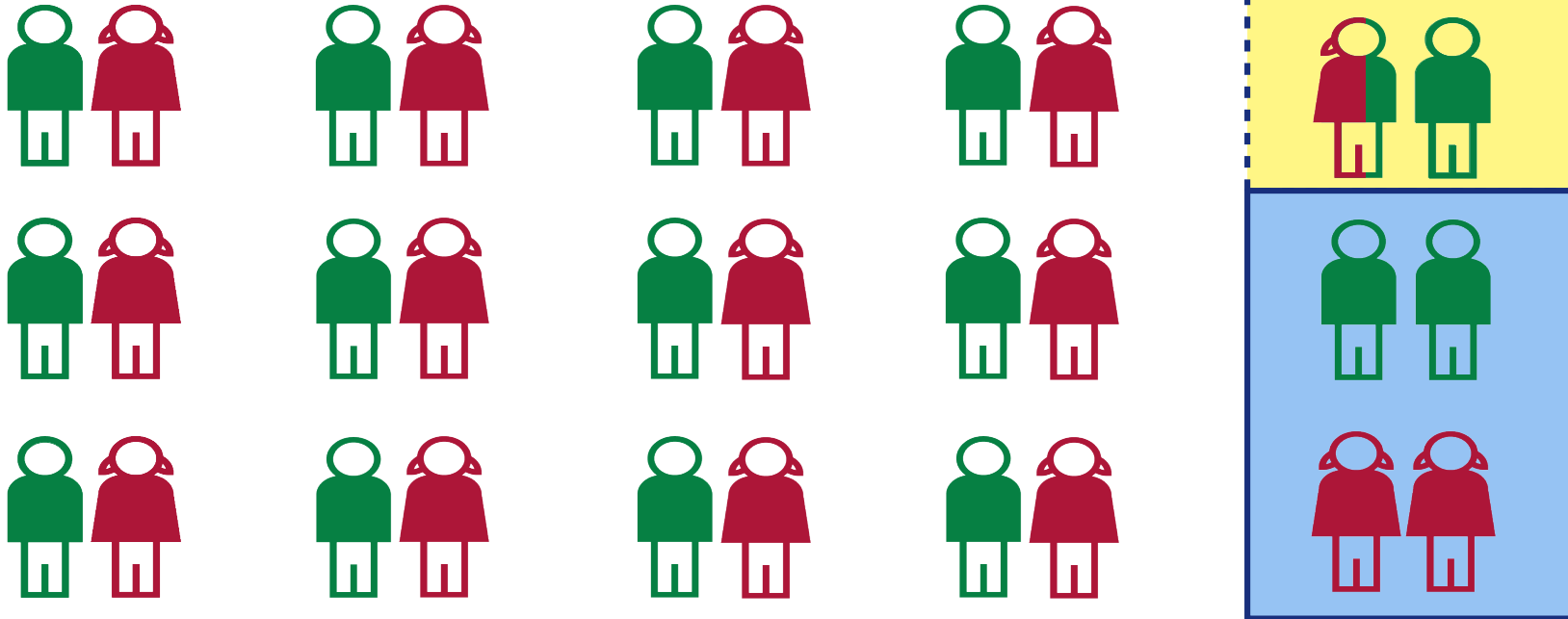
→ **How to overcome this difficulty to take advantage of censuses ?**
(large samples, annual collection of data,...)

30 persons, 15 couples

Among which 6 persons dans 3 SSC → **apparently 20% SSC**

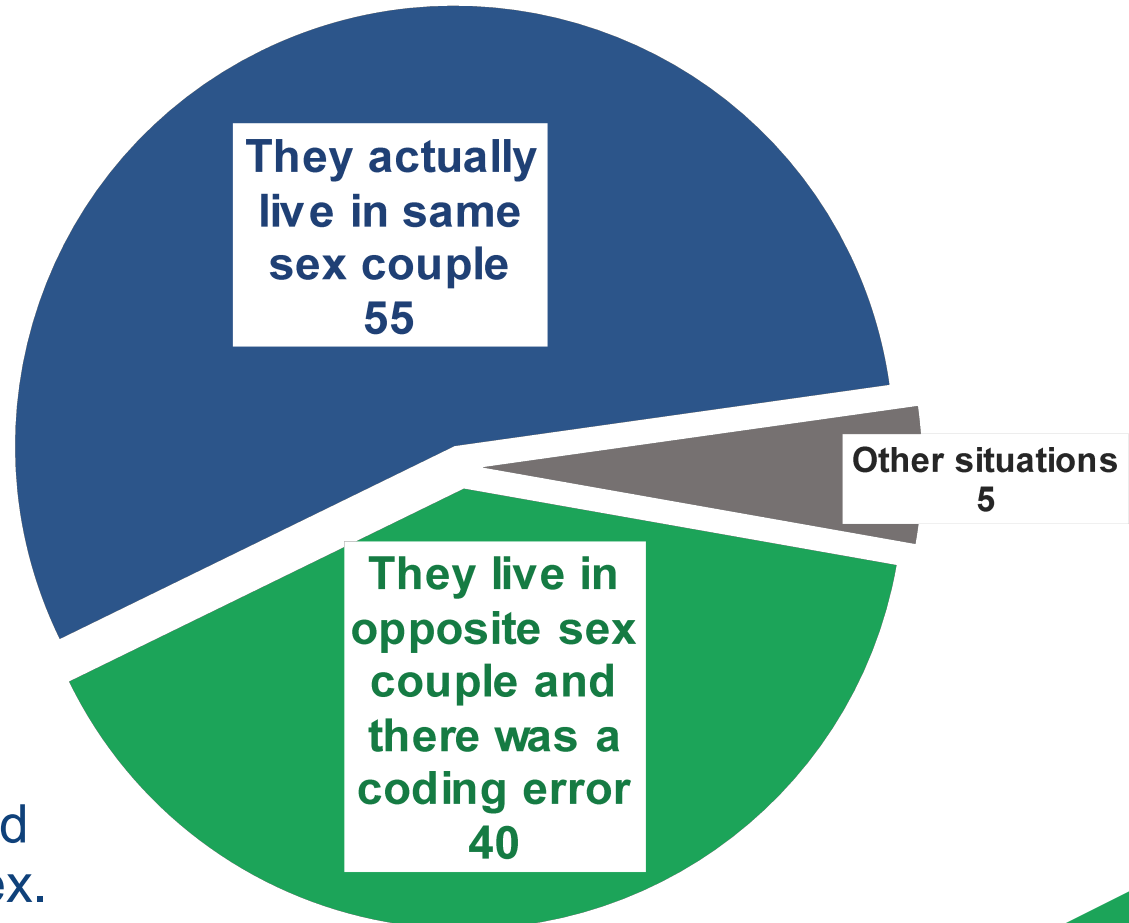


30 persons, 15 couples, 1 coding error (3,3%)
Among which 4 persons in 2 actual SSC → 13% SSC



Family and housing survey 2011 :

When according to census information, persons live in same-sex couple, once checked with specific survey information...



This is consistent with 0.6 % of SSC and 0.17 % of individual coding errors on sex.

1) Information redundancy and cross-checking

- Canadian census (2001)
- American census (2020)
- French Family and household survey 2011

Canadian census question : **Person 2 is...**

- 1: Opposite-sex husband or wife of Person 1
- 2: Opposite-sex common-law partner of Person 1
- 3: Same-sex married spouse of Person 1
- 4: Same-sex common-law partner of Person 1 [...]

2) Matching with administrative data

- American census (2010)
- French permanent demographic sample (EDP)

EDP : coding errors concerns 0.17 % of persons en couple, but 22 % of those apparently in SSC

3) “Statistical validation” using first names

- American census (2010)
- French permanent demographic sample (EDP)

Almost systematic match between supposed sex on the basis of first names and sex recorded in administrative data.

Broad principle : associate to each first name an indicator of « femaleness », proportion of women bearing that first name

- Two sources :
 - First names from vital registry since 1900 (exhaustive for births occurring in France)
 - Completed with first names recorded in the census (for first names of foreign born individuals)
- Tests lead to select an indicator combining different information :

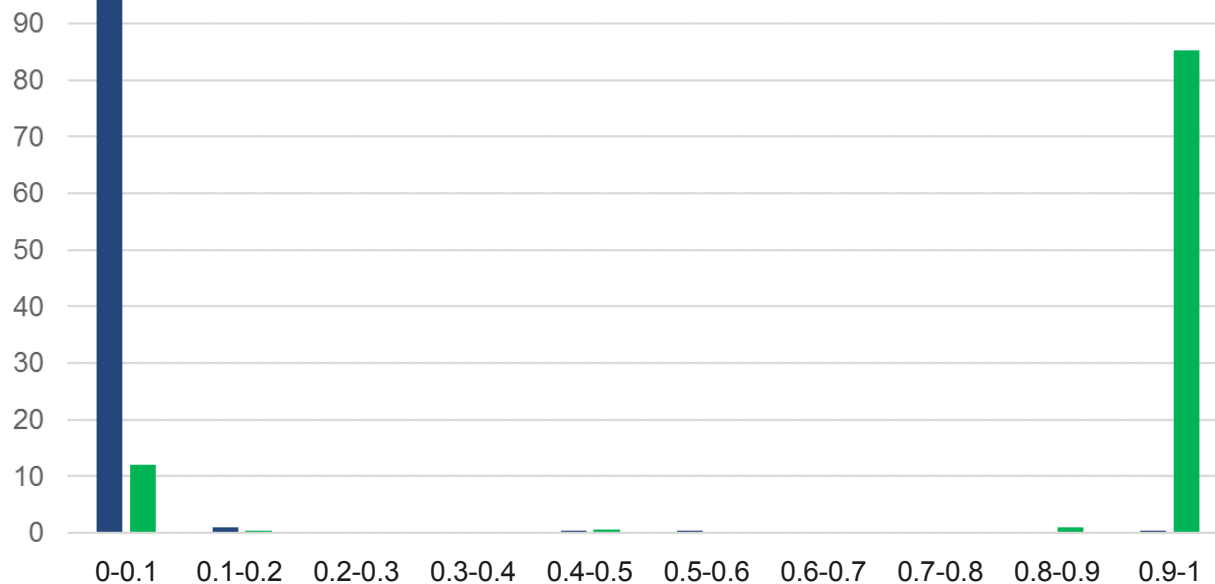
For example Marie-Héloïse was born in 1985 and she answered to the 2017 annual census data collection. According to her answers, she lives in couple with a woman.

Which indicator will she be assigned ?

If the number of « Marie-Héloïse » born in 1985 in the vital registry is sufficient (more than 10), we take the proportion of women among them. If there are too few, conditions are released. In that example we could take a far more less specific indicator : the proportion of women among the « Marie » born since 1900.

Tested on the Permanent demographic sample (EDP), this technique proved to be highly effective in identifying coding errors

Percent
among all persons
living in couple



Indicator of coding error from the dictionary :
proportion of persons of a different sex among those bearing the
same first name

■ **No error :**

**Match between sex from the census and
from administrative data (RNIPP) – 99.83%**

■ **Error :**

**Mismatch between sex from the census and
from administrative data (RNIPP) – 0.17%**

Among persons apparently living in SSC	
N	11 349
Actual errors	2 059
Specificity (en %)	98
Sensitivity (en %)	95
Undue corrections	147
Valid corrections	1 951

Online data collection

Procedure tested on EDP gives satisfying results and is applied

Paper collection

First names are handwritten, then scanned : lower level of quality & risk of undue corrections

Indicator adapted but that leads frequently to intermediate values (close to 50% woman / 50% men).

In those cases, use of additional variables

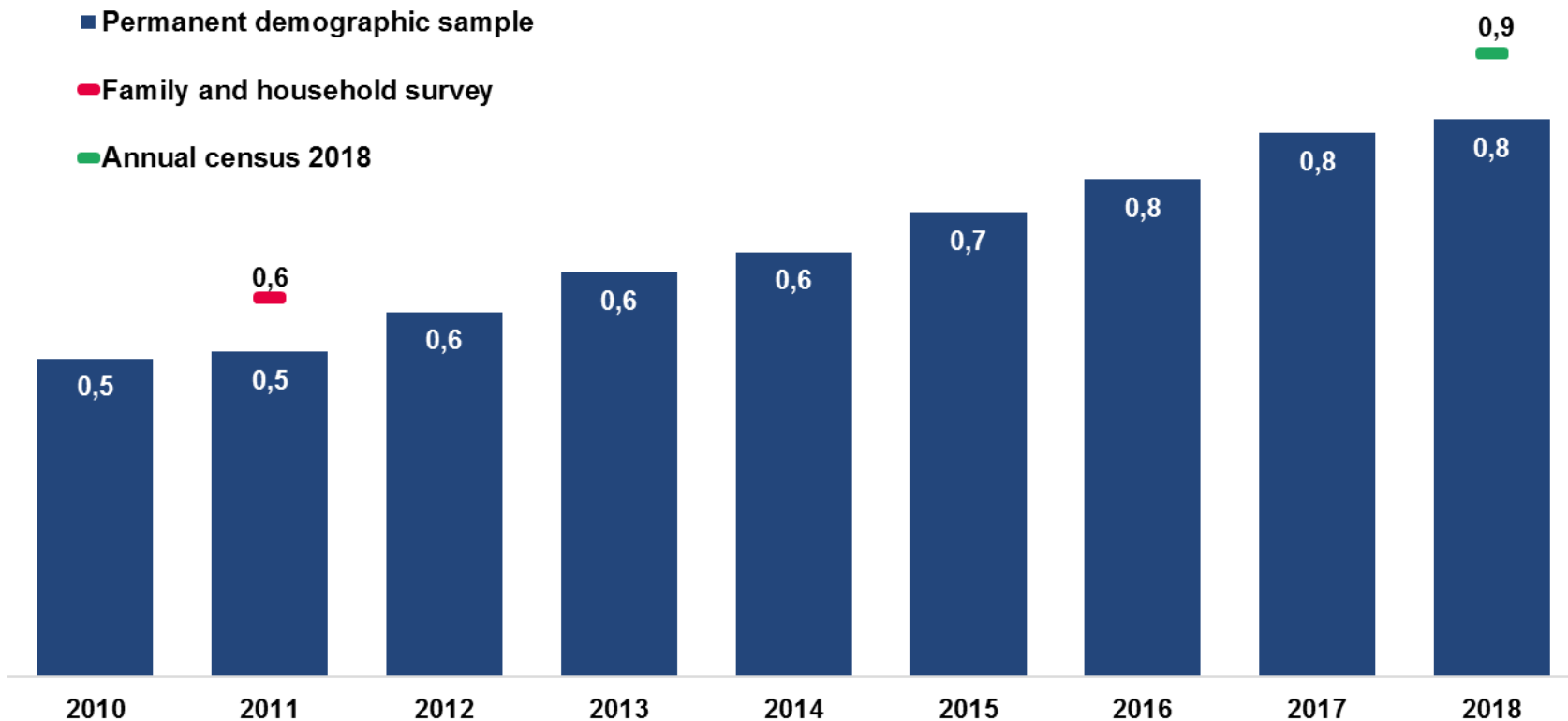
For example, between 2 men apparently in SSC, a coding error is more likely for the one living with children than for the one living with a partner (not a spouse) and without children.

Important limit, but recent developments will reduce its importance :

➤ online data collection

➤ of coding errors with the new questionnaire implemented in 2018

Estimate of the proportion of actual same-sex couples in France (as a per cent of persons cohabiting with a partner)



Sources :

Permanent demographic sample 2010-2018, Insee

Family and household survey 2011, Insee

Annual census 2018, Insee.

Implemented for research purposes

In the 2017 and 2018 census annual data collections

New results soon to be published on same sex couples in France and their characteristics in 2018

Future broader implementation in the census

Correction of coding errors in early stages of census data editing

Correction will affect all available census data

Thank you for your attention !

- For more details, report to working paper (in french) :

ALGAVA E., HALLEPEE S., 2018, « *Estimer les effectifs de couples de personnes de même sexe au recensement : expérimentation d'une solution de validation du sexe par le prénom* », document de travail F1807, Insee

- elisabeth.algava@insee.fr