

School of Political and Social Sciences (PSAD)  
Advanced Master in Quantitative Methods in the Social Sciences

# **Understanding fertility behaviour in Belgium: merging administrative and survey data**

By :

**Armel LEKEUFACK**

Promoters :

**Ester RIZZI, UCLouvain**

**Ingrid SCHOCKAERT, Statistics Flanders**

# **Presentation plan**

**1. Introduction**

**2. Data**

**3. Methods**

**4. Preliminary results**

## 1.1. Background

- ❑ The demographic situation in Belgium: an ageing population and falling birth rates (Vodisek, 2023), changes in birth behaviour (Goldscheider et al., 2015) and a decline in the number of children (Adsera, 2011).
- ❑ A complex link between fertility behaviour and religion that is increasingly difficult to perceive given the multitude of databases and their partial nature..
- ❑ The need to develop techniques that make it possible to combine several sources of data that are not observed in the same survey.

## 1.2. Problem

- ❑ The issue, which was thought to have been settled, is re-emerging in public debates (Ghia, 2018; Honoré, 2021), especially with regard to family and birth policies and the demographic situation in the Western and Belgian contexts.
- ❑ Difficulty in perceiving this role for a number of reasons: the phenomena are less and less observed simultaneously during surveys, the issue of data protection and the decline in voluntary participation in surveys.
- ❑ Question: how does religion affect the arrival of a second child for women who already have one?

## 1.3. Objectives

### □ General objective

- To analyse fertility behaviour in Belgium in relation to religion by merging data from the National Register (NR) and the European Social Survey (ESS).

### □ Specific objectives :

- to carry out a statistical matching of data from the 2011 NR and waves 5 and 6 of the ESS ;
- analyse the influence of religiosity on the probability of having a second child for women who already have one.

## 1.3. Research hypotheses

### □ General hypothesis

- Merging RN and ESS data provides a better understanding of women's fertility behaviour in Belgium by combining the territorial precision of the RN with the attitudinal data of the ESS.

### □ Specific hypotheses:

- the fusion of NR and ESS data offers excellent results for understanding the relationship between fertility and religion.
- the degree of religiosity has a positive influence on the probability of having a second child among women who already have one.

## 2.1. ESS

- ❑ European survey carried out every 2 years since 2002 in 38 countries (11 waves available and 12th in progress).
- ❑ 2 waves of the ESS European survey: 2010 (ESS5) and 2012 (ESS6) were used.
- ❑ ESS is a benchmark survey for attitudinal data in Europe(Peri et al., 2023), including degrees of religiosity.

## 2.2. NR

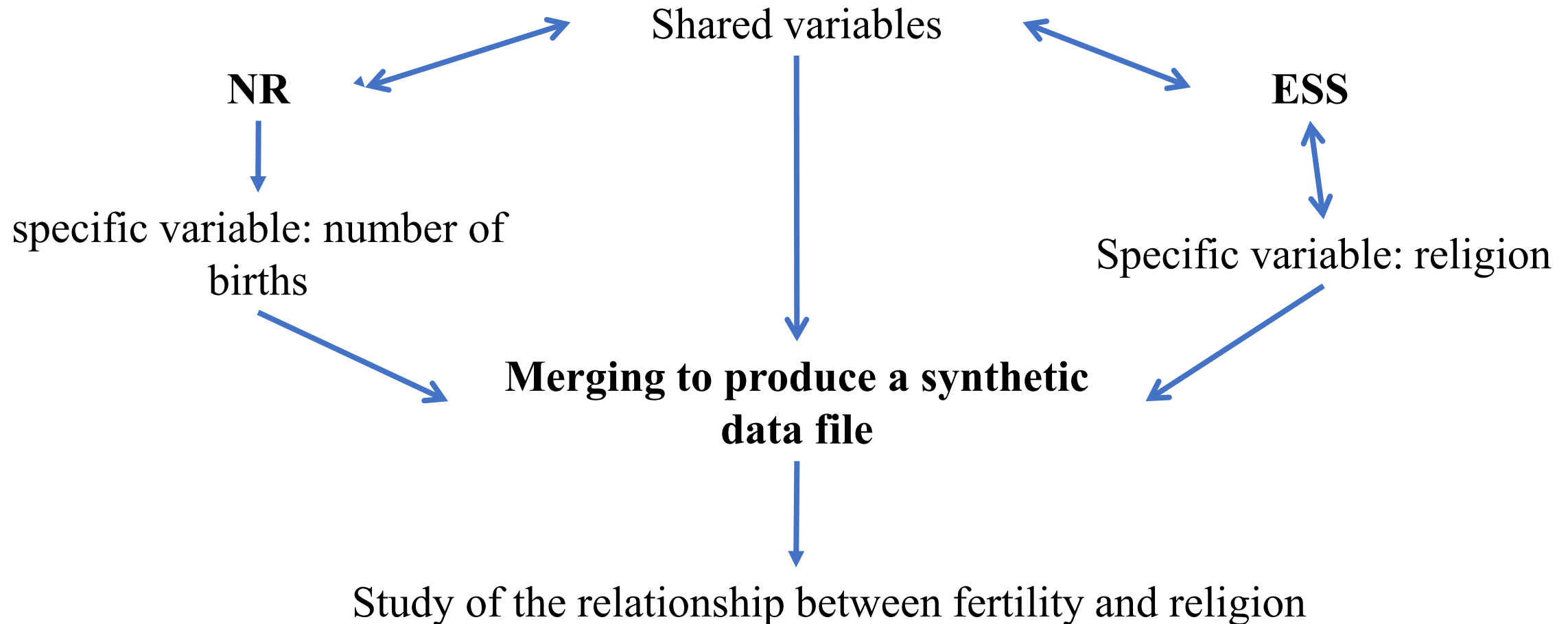
- ❑ Register data as an integration of several data sources
  - Variables contained in CENSUS 2011 and stock 2011
  - Data file containing births since 1998 for women in Belgium
  - Integration by means of a unique identifier of the 3 data sources to obtain a single file of data on women over 18 in Belgium as NR data



### 2.3. Pourquoi ESS5, ESS6, NR and 2011

- ❑ Firstly, we are interested in religion in ESS and births in NR
- ❑ The ESS is based on the NR. Waves 5 and 6 cover the whole of Belgium with a response rate of over 50%, unlike the other waves (in 2020, the sampling plan was based on only 78 large cities for a response rate of only 39.19%).
- ❑ The two waves were combined to increase the number of respondents.
- ❑ This choice of 2011 was made to limit the effects of COVID19 and eco-anxiety, particularly among the younger generations.

### 3.1. Statistical matching: General outline of the method (based on D'Orazio et al., 2006)



### 3.2. Statistical matching: principle

- ❑ There are two databases A and B. Database A contains the specific variable Y and database B the specific variable Z. The set has a series of variables X which are common to both files and come from the same population. The main objective of statistical merging is to create a synthetic file containing information on (X, Y, Z) using the two original files as a bridge.
- ❑ **Approach** : Base on D'Orazio et al (2006), using the StatMatch package in its latest version 1.4.1 on R (D'Orazio, 2022) .

### 3.3. Statistical matching : steps

1. Choice of specific variables: Degree of religiosity (ESS) and Number of births (RN)

2. Harmonisation of variables: identification of common variables and harmonisation

3. Choice of merging variables: Analysis of the distributions of pairs of common variables, Analysis of the predictive capacity between Y and X in A and Z and X in B, Choice of merging variables.

4. Statistical merging of ESS and RN: Two options: (a) ESS as donor and RN as recipient, (b) RN as donor and ESS as recipient.

5. Merge quality: Comparison of marginal distributions, Dissimilarity analysis (TVD)

### 3.4. Relationship between fertility and religion

*Under development...*

## 4.1. Liste des variables communes

Table I : Variable characteristics

<b>Variables</b>	<b>Modalités de la variable</b>
<b>Groupe d'âge</b>	1. 18 à 25 ans, 2. 26 à 35 ans, 3. 36 à 45 ans, 4. 46 à 55 ans, 5. 56 à 65 ans, 6. 66 ans et plus.
<b>Pays de naissance</b>	1. Belgique, 2. hors Belgique.
<b>Nationalité</b>	1. Belge, 2. Non belge.
<b>Occupation</b>	1. travaille, 2. ne travaille pas.
<b>Niveau d'éducation</b>	1. aucun, 2. primaire, 3. secondaire, 4. supérieur.
<b>Statut marital</b>	1. femmes en union légale, 2. femmes qui ne sont pas engagées dans une union, 3. femmes qui n'ont jamais été engagées dans une union légalement reconnue.
<b>Province de résidence</b>	1. Flandre Occidentale, 2. Flandre Orientale, 3. Anvers, 4. Limbourg, 5. Brabant Flamand, 6. Brabant Wallon, 7. Hainaut, 8. Namur, 9. Liège, 10. Luxembourg, 11. Bruxelles-capitale.

## 4.2. Comparison of distributions of common variable pairs

Table II: similarity and dissimilarity of distributions of common variables

Variables	Coef. de Bhatt.	TVD
Groupe d'âge	0,999	0,05
Pays de naissance	0,999	0,03
Nationalité	0,998	0,04
Occupation	0,950	0,27
Plus haut niveau d'éducation	0,992	0,12
Statut marital	0,990	0,12
Province de résidence	0,998	0,04

- Exclusion de la variable occupation à cause d'une différence considérée comme étant trop grande (total variation distance (TVD) = 0,27)

### 4.3. Matching variables (1/2)

Table III: Linear regression anova for the choice of the best predictors for the degree of religiosity

<b>Dependent variable: degree of religiosity</b>					
	<i>Df</i>	<i>Sum sq</i>	<i>Mean sq</i>	<i>fvalue</i>	<i>pvalue</i>
Âge	5	1393	278,61	13,9655	2,070e-13 ***
Niveau d'éducation	3	171	57,07	2,8609	0,035683 *
Statut marital	2	235	117,53	5,8912	0,002819 **
Pays de naissance	1	596	596,11	29,8807	5,254e-08 ***

☐ Variables excluded: nationality, occupation, province of residence.



## 4.5. Matching variables (2/2)

Table IV: Linear regression anova for the choice of the best predictors for births

<b>Dependent variable: number of births</b>					
	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Âge	5	77506464	15501293	373375,9	< 2,2e-16 ***
Niveau d'éducation	3	2174643	724881	17460,0	< 2,2e-16 ***
Statut marital	2	41436155	20718078	499031,4	< 2,2e-16 ***
Province de résidence	10	379046	37905	913,0	< 2,2e-16 ***
Pays de naissance	1	375936	375936	9055,1	< 2,2e-16 ***

□ Best predictors : age, level of education, marital status and country of birth were used as **MATCHING VARIABLES**

## 4.6. Validity check : analysis of the marginal distribution

Tableau IV : analyse des distributions marginales des variables spécifiques dans le donneur et le receveur

Processus 1			Processus 2		
Receveur : ESS – donneur : RN			Receveur : RN – donneur : ESS		
Nombre de naissances	dist_don	dist_final	Degrés de religiosité	dist_don	dist_final
0	0	0,03	0	0	-6,21
1	0	0,80	1	0	3,01
2	0	-1,91	2	0	2,09
3	0	-0,05	3	0	0,83
4	0	1,19	4	0	0,19
5	0	1,20	5	0	0,06

#### 4.7. Validity check : total distance of variation

- ❑ The calculation of the total distance of variation between the distributions (number of births in process 1 and degrees of religiosity in process 2) shows a very high overall dissimilarity in process 2 (ESS as receiver and RN as donor). This dissimilarity is in fact 112.26% compared with a reference distribution in the donor (RN), highlighting a major difference between the observed and expected distributions.
- ❑ For process 2, this dissimilarity was 11%. This value shows a moderate dissimilarity, proving that the two distributions are close, although there are notable differences.

#### 4.7. Validity check : total distance of variation

- ❑ A comparison of the marginal distributions and dissimilarities shows that the process that considers RN as the receiver and ESS as the donor produces better results than the one that considers the reverse.
  
- ❑ The synthetic data file obtained from process I is the one used in the next stage of the case study, i.e. the impact of the degree of religiosity on the probability of having or not having another child for women who already had one.
  
- ❑ This is a database containing the degree of religiosity, the number of children and the 7 initial common variables.

## 4.8. Relationship between fertility and religion

*Data currently being analysed...*

*Thank you for your attention!!!*