Documentation for the International Sibsize and Educational Attainment Database (ISEAD)

19 January 2018

Version 1.0

This research has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme under grant agreement No 681546 (FAMSIZEMATTERS).

1. Introduction

The International Sibsize and Educational Attainment Database (ISEAD) consists of two parts:

1) Data on sibling size and educational attainment for 166 birth cohorts in 26countries aggregated from a pooled dataset of 114 surveys.
2) Stata code to produce from scratch the pooled dataset on which the aggregated data are based; this code allows others to replicate our analyses and to run their own related analysis.

Sibling size in our database indicates family size from the child's perspective, which is different from the family size of women (Fahey, 2017; Präg, Choi, & Monden, 2018; Preston, 1976).

We can only share aggregated data and code to reconstruct the pooled database from scratch. We are not allowed to share the individual surveys, but we provide links to the data archives/providers where you can obtain the raw data.

2. Funding

This research was supported by a European Research Council Consolidator Grant to Christiaan Monden for the FAMSIZEMATTERS project (proposal number 681546).

3. Criteria for inclusion of surveys

We collected surveys that contain information on respondent's number of siblings (and other aspects of sibling configuration like sex composition, and birth order, if available), respondent's own education & occupation, and also information on their parents' socioeconomic status measured by education and/or occupation.

We considered the following criteria in the selection of surveys:

1) The survey sample should represent a national population of adults. We are flexible about the age range of those adults, but in principle we consider adults who have completed their formal education (e.g., age 27 or over). This means we do not include national samples from adolescent or child studies.
2) Initially our database covers countries with low fertility. We define low fertility as below-replacement fertility, but this definition is rather loosely applied in practice; our database includes all countries in Europe (Western and Eastern), East Asia (China, Japan, South Korea and Taiwan) and OECD countries in other regions (United States, Canada, Australia) with representative survey samples available.
3) We limit our time window to data collected in the second half of the past century and later (i.e. surveys collected from 1950 onwards). .
4) The surveys has to include information on the respondent's educational attainment in levels, qualifications or number of years (not ability tests or grades), parents' educational attainment and/or occupation and the respondents' number of siblings.

4. Search strategy

We tried to find all relevant surveys by following these four steps:

1) For each survey in the International Stratification and Mobility File (a collection by Harry Ganzeboom, Donald Treiman, and Elizabeth Stephenson, see http://www.harryganzeboom.nl/ismf/index.htm) we checked whether there was information on sibling size
2) We searched the following data archives for relevant surveys:
   i. ICPSR https://www.icpsr.umich.edu/icpsrweb/ICPSR/
   ii. CESSDA http://cessda.net/
   iii. GESIS http://www.gesis.org/en/services/data-analysis/data-archive-service/
3) We checked large international survey programmes for information on siblings: ISSP, ESS, SHARE, FFS, GGP, EASS.
4) Followed references in cross-nationally comparative social mobility studies, in particular Breen's 2004 Social Mobility in Europe (Breen, 2004).
5) For countries where we couldn't identify surveys using 1-4 we searched the national data archive.

5. Current state of collection

As of 19/1/2018 (version 1.0), we have collected 114 samples from 26 countries.

6. Variables in the pooled data

The key variables of our pooled micro database are: Respondent's number of siblings; Respondent's own education; Respondents' parent's education; Birth year (and birth cohort); Age at the time of interview; Survey year; Sex.

We also collect occupational status of one's own and parent's when they are available. When it comes to sibling number, we provide separate variables for brothers and sisters if such information is available. Birth order is also included where it is asked. In the following discussion, we provide below a brief outline of how each of these variables was constructed. The aggregated data by birth cohort is derived from this pooled database.

In measuring respondent and parental education, we used two specifications: (1) Educational years and (2) educational categories. Educational years captures the nominal years of education a respondent or his or her parents had completed at the time of the survey. While some surveys measured this directly, others asked respondents to report the highest level of education they or their parents had completed. In these instances, we used country-specific correspondence tables provided by UNESCO (Link: http://uis.unesco.org/en/isced-mappings), which map country-specific educational levels to the respective number of years typically

required to attain that level within that country. For the very few countries where no UNESCO mapping was available, we used education recode guides provided by Harry Ganzeboom (see: http://www.harryganzeboom.nl/ismf/ismf.htm).

To code educational categories, we used the three-digit 2011 International Standard Classification of Education (ISCED) classification. At the its broadest level, ISCED classifies educational attainment into 9 categories: (0) Early childhood education; (1) Primary education; (2) Lower secondary education; (3) Upper secondary education; (4) Post-secondary non-tertiary education; (5) Short-cycle tertiary education; (6) Bachelor or equivalent; (7) Master or equivalent; (8) Doctoral or equivalent. Within each of these broad categories, there are varying numbers of sub-categories that further differentiate educational attainment. While a select few surveys measured educational attainment using the 2011 ISCED classification, the majority of surveys utilized in our dataset measured educational attainment using some other country-specific scheme. To translate these disparate codings into the ISCED 2011 classification, we again referred to ISCED 2011 mapping tables provided by UNESCO (Link: http://uis.unesco.org/en/isced-mappings). For the very few countries where no UNESCO mapping was available, we also again used education recode guides provided by Ganzeboom (see: http://www.harryganzeboom.nl/ismf/ismf.htm).

Sibship size was constructed as the total number of sisters and brothers a given respondent has ever had, not including the respondent herself. Many surveys did not specify whether respondents should count half- or non-biological siblings such as step-siblings and adopted siblings, while others clarified that siblings should include the number of full, half, and non-biological siblings. Since we did not identify any single survey that asked for the number of biological siblings specifically, we assumed that all values of the number of siblings represented sibship size inclusive of all types of siblings. To minimize an arbitrary variation in the upper bound of sibship size across surveys, we cap all sibship size values higher than 10 at 10 before aggregating the data.

Occupation of the respondent and his or her parents was coded, when possible, using the International Standard Classification of Occupations (ISCO) 2008 scheme. For older surveys, occupation was coded using the 1988 ISCO scheme (with a flag for ISCO scheme). As educational attainment (rather than occupational attainment) was the central focus of this project, many of the surveys included in this dataset did not include an ISCO measure of occupation, or measured occupation using some country-specific scheme. In these instances, occupation was either left as missing or was included in our dataset using the raw country-specific measure.

Lastly, we measured age and sex using self-reported answers provided in each surveys. Birth cohort was constructed by subtracting the respondent's age from the survey year.

7. Variables in ISEAD aggregated data

Country
Birth cohort
Average sibsize
Average education
Average parental education
N

8. Stata code

We assigned an acronym to each dataset we used by combining the three-letter ISO country code and the four-digit survey year. This acronym is also the name of the do file of each dataset. To replicate, please download the corresponding do files provided on the ISEAD dataverse.

9. Some notes on the data compilation

We came across several challenges in the data collection process. To begin with, many major surveys do not include information on siblings. Even when they do, there can be many issues: some surveys only ask children or adolescent about their siblings, while some only survey the number of siblings within the same household at the time when the survey was done. Moreover, we had to focus on the total number of siblings, because information on sibling configuration, such as number of sisters/brothers and birth order, is not readily available in most surveys we found. Furthermore, there are some gender issues to consider: A few samples collected earlier than 1980s, especially from Japan and the UK, surveyed only men. We address this issue either by analysing men and women separately or by controlling for sex. For parental education, many surveys asked father's but not mother's. Some asked education of a parent without specification. Of course, more recent surveys tend to ask education of both parents. We address this issue by modelling how parental education is measured. When information of both parents is available, we use fathers for consistency.

10.List of surveys

Below is a list of all the surveys we included in ISEAD. We also provide the links to download these surveys. Please note that there are two types of data access: some data can be downloaded directly from the website after signing up, some require formal application process before usage.

Country	Year	Survey Name	Acronym	Access	Note
Australia	2005	Generations and Gender Programme	AUS2005	Link	Wave 1
Belgium	2008	Generations and Gender Programme	BEL2008	Link	Wave 1
Bulgaria	1993	Social Stratification in Eastern Europe After 1989: General Population Survey	BGR1993	Link	
Bulgaria	2004	Generations and Gender Programme	BGR2004	Link	Wave 1
Canada	1982	Canadian Class Structure and Class Consciousness Survey	CAN1982	Link	
Canada	1995	Canadian General Social Survey	CAN1995	Link	
Canada	2000	Canadian General Social Survey	CAN2000	Link	
China	1993	China Housing Survey	CHN1993	Link	
China	1998	Chinese Longitudinal Healthy Longevity Survey (CLHLS)	CHN1998	Link	Wave 1
China	2006	East Asian Social Survey: Families in East Asia	CHN2006	Link	
China	2010	China Family Panel Studies	CHN2010	Link	Wave 1
Czech Republic	1993	Social Stratification in Eastern Europe After 1989: General Population Survey	CZE1993	Link	
Czech Republic	2004	Generations and Gender Programme	CZE2004	Link	Wave 1
Estonia	2004	Generations and Gender Programme	EST2004	Link	Wave 1
France	1985	Training and Qualifying Survey Professional (Formation, Qualification Professionnelle)	FRA1985	Link	
France	1993	Training and Qualifying Survey Professional (Formation, Qualification Professionnelle)	FRA1993	Link	
France	2003	Training and Qualifying Survey Professional (Formation, Qualification Professionnelle)	FRA2003	Link	
France	2005	Generations and Gender Programme	FRA2005	Link	Wave 1

Georgia	2006	Generations and Gender Programme	GEO2006	Link	Wave 1
Germany - East	1990-2015	The German Socio-Economic Panel	EDEU1981	Link	
Germany - West	1984-2015	The German Socio-Economic Panel	WDEU1981	Link	
Germany	2005	Generations and Gender Programme	DEU2005	Link	Wave 1
Hungary	1983	Hungarian Social Mobility and Life History Survey	HUN1983		
Hungary	1986	International Social Survey Panel (ISSP)	HUN1986	Link	
Hungary	1993	Social Stratification in Eastern Europe After 1989: General Population Survey	HUN1993	Link	
Hungary	2004	Generations and Gender Programme	HUN2004	Link	Wave 1
Italy	2003	Generations and Gender Programme	ITA2003	Link	Wave 1
Japan	1995	Japanese Social Stratification and Mobility	JPN1995	Link	Data in Japanese
Japan	2000	Japanese General Social Survey	JPN2000_1	Link	
Japan	2000	Japan 2000 National Survey on Family and Economic Conditions	JPN2000_2	Link	
Japan	2001	Japanese General Social Survey	JPN2001	Link	
Japan	2002	Japanese General Social Survey	JPN2002	Link	
Japan	2005	Japanese General Social Survey	JPN2005_1	Link	
Japan	2005	Japanese Social Stratification and Mobility	JPN2005_2	Link	Data in Japanese
Japan	2006	Japanese General Social Survey	JPN2006_1	Link	
Japan	2006	East Asian Social Survey	JPN2006_2	Link	
Japan	2008	Japanese General Social Survey	JPN2008_1	Link	
Japan	2008	National Family Research of Japan	JPN2008_2	Link	
Japan	2009	Japan 2009 National Survey on Family and Economic Conditions	JPN2009	Link	
Japan	2010	Japanese General Social Survey	JPN2010	Link	
Japan	2012	Japanese General Social Survey	JPN2012	Link	
Lithuania	2006	Generations and Gender Programme	LTU2006	Link	Wave 1
Netherlands	1992	Family Survey of the Dutch Population (Family-enquete Nederlandse Bevolking)	NLD1992	Link	
Netherlands	1998	Family Survey of the Dutch Population	NLD1998	Link	

		(Family-enquete Nederlandse Bevolking)			
Netherlands	2000	Family Survey of the Dutch Population (Family-enquete Nederlandse Bevolking)	NLD2000	Link	Data in Dutch
Netherlands	2002	Generations and Gender Programme	NLD2002	Link	Wave 1
Netherlands	2002	Netherland Kinship Panel Survey	NLD2002_2	Link	Wave 1
Netherlands	2003	Family Survey of the Dutch Population (Family-enquete Nederlandse Bevolking)	NLD2003	Link	Data in Dutch
Norway	2007	Generations and Gender Programme	NOR2007	Link	Wave 1
Poland	1988	POLPAN (Polish Panel Survey)	POL1988	Link	1988 Wave
Poland	1993	Social Stratification in Eastern Europe After 1989: General Population Survey	POL1993	Link	
Poland	1997	Polish General Social Survey	POL1997	Link	
Poland	1999	Polish General Social Survey	POL1999	Link	
Poland	2002	Polish General Social Survey	POL2002	Link	
Poland	2010	Generations and Gender Programme	POL2010	Link	Wave 1
Romania	2005	Generations and Gender Programme	ROU2005	Link	Wave 1
Russia	1992	Russian General Social Survey	RUS1992	Link	
Russia	1993	Social Stratification in Eastern Europe After 1989: General Population Survey	RUS1993	Link	
Russia	2004	Generations and Gender Programme	RUS2004	Link	Wave 1
South Korea	1990	Inequality and Equity Survey	KOR1990	Link	Data in Korean
South Korea	2004	Korean General Social Survey	KOR2004	Link	
South Korea	2006	East Asian Social Survey	KOR2006_1	Link	
South Korea	2006	Korean General Social Survey	KOR2006_2	Link	
South Korea	2008-2011	Education and Social Mobility Survey	KOR2011	Link	Data in Korean
Spain	2013	Occupational Prestige and Social Structure (Prestigio Ocupacional Y Estructura Social)	ESP2013	Link	
Sweden	2012	Generations and Gender Programme	SWE2012	Link	Wave 1
Taiwan	1997	Taiwan Social Change Survey	TWN1997_1	Link	

Taiwan	1997	Taiwan Social Change Survey	TWN1997_2	Link	
Taiwan	1999	Panel Study of Chinese Family Dynamics	TWN1999	Link	Data in Chinese; 1999 wave
Taiwan	2002	Taiwan Social Change Survey	TWN2002	Link	
Taiwan	2004	Taiwan Social Change Survey	TWN2004	Link	
Taiwan	2006	East Asian Social Survey	TWN2006	Link	
Taiwan	2007	Taiwan Social Change Survey	TWN2007	Link	
Taiwan	2012	Taiwan Social Change Survey	TWN2012	Link	
Taiwan	2014	Taiwan Social Change Survey	TWN2014	Link	
Taiwan	2015	Taiwan Social Change Survey	TWN2015	Link	
United Kingdom	1972	Oxford Social Mobility Inquiry	GBR1972	Link	
United Kingdom	1980	UK National Heights and Weights Survey	GBR1980	Link	
United Kingdom	1986	International Social Survey Panel (ISSP)	GBR1986	Link	
United Kingdom	1990	UK National Survey of Sexual Attitudes and Lifestyle	GBR1990	Link	
United Kingdom	2000	UK National Survey of Sexual Attitudes and Lifestyle	GBR2000	Link	
United Kingdom	2009-2011	UK Household Longitudinal Study (Understanding Society)	GBR2009	Link	Wave 1
United Kingdom	2010	UK National Survey of Sexual Attitudes and Lifestyle	GBR2010	Link	
United States	1955	Growth of American Families	USA1955	Link	
United States	1972-2014	General Social Survey	USA1972 - USA2014	Link	

11. References

Breen, R. (Ed.) (2004). *Social Mobility in Europe*. Oxford: Oxford University Press.

Fahey, T. (2017). The Sibsize Revolution and Social Disparities in Children's Family Contexts in the United States, 1940–2012. *Demography, 54*(3), 813-834. doi:10.1007/s13524-017-0568-0

Präg, P., Choi, S., & Monden, C. W. S. (2018). Sibship Size in Low-Fertility Countries over the Twentieth Century: Declining Social Disparities. *Socarxiv*. doi:10.17605/OSF.IO/ZVB9C

Preston, S. H. (1976). Family Sizes of Children and Family Sizes of Women. *Demography, 13*(1), 105-114. doi:10.2307/2060423