

18 Europe

PAU Census Microdata Samples Project

Nikolai Botev

Introduction

Since 1992, the Population Activities Unit (PAU) of the Economic Commission for Europe (ECE), in cooperation with the United Nations Population Fund (UNFPA) and the U.S. National Institute on Aging (NIA), has been coordinating a project that resulted in the creation of a collection of cross-nationally comparable census microdata samples. As of June 2000, this collection covered fifteen countries in Europe and North America. Data processing has been completed for most of these fifteen countries. Preliminary contacts on the terms of participation of several new countries have taken place. All samples currently in the collection are based on the 1990-round of national population and housing censuses.

Although the main objective of the project is to study the social and economic conditions of older persons, the samples are designed to allow research on a wide range of issues related to aging, as well as on other social phenomena. A common set of nomenclatures and classifications, derived on the basis of a

Nikolai Botev is the Project Manager for the Population Activities Unit (PAU) of the United Nation's Economic Commission on Europe. The PAU works closely with national population institutes, national statistical offices and individual scholars in the areas of international migration and migrants; population of aging and older persons; and fertility, family and reproductive health.

For more information on this international census project, see the PAU web site at: http://www.unece.org/ead/pau/p_home1.htm.

study of census data comparability in Europe and North America, was adopted as standards for harmonization and recoding.

Source Material

Census microdata were obtained directly from the National Statistical Offices (NSOs) of the participating countries. The samples were drawn by the PAU or NSOs from the complete census files, thus the universes they represent are all persons and housing units in the participating countries. Most of the meta data and documentation related to the samples was obtained directly from the NSOs. Some documentation was made available by the ECE's Statistical Division, which had carried out an independent study of the national practices during the 1990 round of censuses.

Procedural History

The recommendations regarding the design and size of the samples prepared for the project envisaged: (1) drawing individual-based samples of about one million persons; (2) progressive oversampling with age in order to ensure sufficient presentation of various categories of older people; and (3) retaining information on all persons co-residing in the sampled individual's dwelling unit. Most countries have drawn their samples in accordance with these principles. Some countries (specifically Estonia, Finland, Latvia and Lithuania) adhered to earlier recommendations and sampled only the population over age 50 (Estonia, Latvia and Lithuania provided the entire population over age 50, while Finland sampled it with progressive over-sampling). Several countries provided samples that had not been drawn specially for this project, and cover the entire population without over-sampling.

The processing of the data sets, which included drawing of the samples from the complete census files (when requested by the National Statistical Offices), cleaning (where necessary), and standardization/harmonization, was performed by the PAU and every effort was made to ensure quality and comparability.

Electronic Formats

The main medium for data distribution are CD-ROMs. The samples are distributed by the PAU, as the data producer, as SAS transport data files. ICPSR/NACDA, as the collection's main distributor, produces also an ASCII version of the data files, and includes separate files of SAS and SPSS data definition statements to describe the ASCII data file.

Variable Availability

Variable availability varies widely from country-to-country. The attached summary tables identify variable availability according to the standards developed by the PAU for housing and population variables, respectively. Countries identified by "Ø" in Tables 18-1 and Table 18-2 are not participating in the project, and microdata for those censuses will not be available to researchers. For purposes of comparison, both types of countries are included in Tables 18-1 and 18-2. In addition to the traditional censuses that were taken in most countries in the ECE region, the project attempts to cover register-based censuses, carried out in the Nordic countries. The information for Finland, where the 1990 census was entirely register based, was compiled from documentation provided by Statistics Finland. The information for Norway and Sweden covers only the data collected through census forms. The 1990 census of Norway combined register-based estimates with information collected through census questionnaires sent to 10 percent of the population over 16 years of age. Sweden combined register-based estimates with a limited questionnaire covering the total population. Hence, while for Norway the questionnaire provides a relatively complete picture of the census data available, for Sweden the information is only partial.

In Tables 18-1 and 18-2, core variables are marked with gray shading in the summary tables and by bold typeface in each table's index. The index defines the columns in each summary table and the major topics for which data were collected (e.g., birth, household structure, fertility, etc.). Within each major topic most countries collected information on several specific items. Listed in the index and tables are items for which at least

5 countries collected data. Core variables were selected from among these items. For example, columns 1.1, 1.2, and 1.3 of the summary table on population data are core variables pertaining to birth. Every participating country collected these data. Column 1.1 corresponds to "Date of Birth," listed as item 1.1 in the index. Column 1.2 is "Sex" and so on. An "X" next to a country's name in column 1.1, means that the country asked a question about "Date of Birth" in its most recent census. Examining the number of "X" marks in each column provides researchers with a summary of the number of countries that asked this question, thus, providing a guide to cross-national data availability. Several marks other than "X" appear in the summary tables. Their meanings are as follows:

Codes for Housing Data

Summary Table 18–1

(Bold type indicates that item is a Core Variable)

1.0 Structure and Building Characteristics

- 1.1 Number of Units
- 1.2 Number of Floors
- 1.3 Location of Dwelling (Floor)
- 1.4 Main Usage (Residential, recreational, commercial, etc.)
- 1.5 Official Usage (Zoning)
- 1.6 Type of Structure Within Which Dwelling is Located
- 1.7 Work / Live Potential
- 1.8 Construction Date**
- 1.9 Renovation / Rehabilitation
- 1.10 Building Materials**
- 1.11 Group Dwelling or Single Family
- 1.12 Dwelling Status (Occupied, vacant, etc.)

2.0 Utility Systems

- 2.1 Water Source Location**
- 2.2 Hot Water**
- 2.3 Sewage System Type**
- 2.4 Electricity Supply
- 2.5 Gas Supply
- 2.6 Heating System Type (Source)
- 2.7 Heating Fuel

3.0 Amenities

- 3.1 Telephone
- 3.2 Elevator
- 3.3 (Flush) Toilet**
- 3.4 Automobile
- 3.5 Bath / Shower**
- 3.6 Kitchen

4.0 Occupants

- 4.1 Household Relationships
- 4.2 Number of Families Within Dwelling Unit
- 4.3 Ownership (Who owns the unit?)**
- 4.4 How did the person come to occupy the unit?
- 4.5 Entry Date in Dwelling
- 4.6 Number of Inhabitants

5.0 Floor Plan

- 5.1 Dwelling Size (m²)**
- 5.2 Number of Rooms**
- 5.3 Room Size
- 5.4 Room Use

Key to Table 18–1 and 18–2

- X** Information collected on census long form only
- x'** Information can be derived, or comes from a register
- E** Extensive information on the topic
- B** Basic information on the topic
- FSQ** Family Studies Questionnaire
- S** Sources of income
- A** Amount of income
- C** Local community
- R** Current residence

Table 18-1. Housing Variable Availability: PAU Census Microdata Project
(for explanation of variable key and notes, see text)

Country	Amenities										Occupancy					Floor Plan				Number not Labeled
	3.1	3.2	3.3	3.4	3.5	3.6	4.1	4.2	4.3	4.4	4.5	5.1	5.2	5.3	5.4					
Ø Austria H		x			x	x	x	x	x	x		x	x			8				
Belgium H	x		x	x	x	x	x	x	x			x	x		x	1				
Bulgaria P&H			x		x	x	x'					x			x	4				
Canada P&H							x						X	X		13				
Czech Republic H		x							x							2				
Czech Republic P&H	x		x	x	x	x	x	x	x	x		x	x	x		7				
England & Wales P&H			X	X	X		X	X	X	X		X	X			6				
Estonia P&H					x		x	x				x	x			2				
Finland*	?	?				x'	x'					x'	x'			*				
Ø France H			x	x	x	x	x	x	x	x						2				
Ø Germany H			x		x	x	x	x	x	x		x	x	x		3				
Hungary H	X				x	x	x'	x	x	X		X	x	x		4				
Israel P&H	x		x	x	x	x		x	x	x		x	x			4				
Italy P&H	x	x			x	x		x	x	x		x	x			2				
Latvia P&H					x	x	x	x	x			x	x			2				
Lithuania P&H					x	x	x	x	x			x	x			2				
Norway H*		x	x	x	x	x	x	x	x			x	x'	x		2				
Poland P&H			x		x	x	x	x	x	x		x	x			6				
Romania P&H			x		x	x	x'	x	x	x		x	x	x		3				
Russian Fed. P&H					x	x	x	x	x	x		x	x			2				
Spain P&H	x				x	x		x	x	x		x	x			5				
Sweden P&H*		x			x	x	x'		x	x		x	x			5				
Switzerland H						x						x	x			4				
Turkey P							x'									8				
United States P&H	X		X	X	X	X	x	x'	x		X			X		32				

Notes for Population Data

Summary Table 18–2

1.0 Birth

- | | |
|------------|--|
| 1.1 | Date of Birth/Age |
| 1.2 | Sex (Gender) |
| 1.3 | Nation of Birth or Foreign Born |
| 1.4 | Region of Birth Within Country |

2.0 Identity

- | | |
|------------|----------------------|
| 2.1 | Ethnic Group |
| 2.2 | Citizenship |
| 2.3 | Mother Tongue |
| 2.4 | Other Languages |
| 2.5 | Religion |

3.0 Household Structure

- | | |
|------------|-----------------------------------|
| 3.1 | Relationship to Key Person |
|------------|-----------------------------------|

4.0 Marital Status

- | | |
|------------|---|
| 4.1 | Marital Status |
| 4.2 | Date of First Marriage |
| 4.3 | Date of Current or Last Marriage |
| 4.4 | Number of Marriage(s) |
| 4.5 | Reason for Marriage to End |
| 4.6 | Cohabitation |

5.0 Fertility and Mortality

- | | |
|------------|---|
| 5.1 | Children Ever Born (Live-births) |
| 5.2 | Children Surviving |

6.0 Education

- | | |
|------------|-------------------------------|
| 6.1 | Highest Level Attained |
| 6.2 | Type |
| 6.3 | Currently Enrolled |
| 6.4 | Field of Study |

7.0 Income

- | | |
|------------|------------------------------------|
| 7.1 | Primary Source of Income |
| 7.2 | Primary Source(s) of Income |

8.0 Economic Activity

- 8.1 Employed (Full time/Part time/Not employed)
- 8.2 Unemployed (Y/N)
- 8.3 Employer Type: Employer, Employee, Self-employed, Family Worker
- 8.4 Type of Business in Which One Works: Sector / Industry
- 8.5 Category of Economic Inactivity
- 8.6 Economically Active One Week Prior to Census Date (Y/N)
- 8.7 Employed One Year Prior to Census Date (Y/N)

9.0 Occupation

- 9.1 Occupation
- 9.2 Employed Full- or Part-time in Main Occupation
- 9.3 Any Other Employment
- 9.4 Commute to Work (Frequency)
- 9.5 Time Spent Commuting
- 9.6 Place of Work
- 9.7 Mode of Transportation

10.0 Residential Mobility

- 10.1 Residence x Years Ago ($0 < x < 10$ years)
- 10.2 Location of Previous Residence
- 10.3 Duration of Current Residence

11.0 Disabilities

- 11.1 Disabilities

12.0 Other Persons Present

- 12.1 Address
- 12.2 Reason
- 12.3 Duration of Presence

13.0 Household Members Absent

- 13.1 Address
- 13.2 Reason
- 13.3 Duration of Absence

14.0 Institutional Population

- 14.1 Institutional Population

Data Access

Beta and pre-release versions of six data sets are already available through NACDA/ICPSR. Table 18-3 summarizes the status of data acquisition, processing, and access conditions for the participating countries.

Table 18-3: PAU Census Micro-Data Project Status of Data Acquisition and Processing for the Participating Countries (listed in order of receipt)

Countries	Sampling		Data Processing Status	Data Release Access Conditions ³	Availability (8/11/01)
	Design ¹	Sample drawn by			
USA	No	1990 PUMS ²	In proc.	general	Oct. 1999
Estonia	Partially	NSO	DONE	general	pre-release
Finland	Partially	NSO	DONE	general	beta
Romania	Yes	NSO	DONE	general	beta
Switzerland	Yes	NSO	DONE	limited	beta
Bulgaria	Yes	PAU	DONE	general	beta
Hungary	Yes	NSO	DONE	limited	July 1999
Czech Republic	Yes	PAU	DONE	general	beta
Latvia	Partially	NSO	DONE ⁴	general	beta
Turkey	No	1990 SIS 5% sample ²	DONE	general	Oct. 1999
Lithuania	Partially	NSO	DONE ⁴	general	Dec. 1999
Russia	No	NSO	DONE	limited	Dec. 1999
Canada	No	1991 PUMFs ²	In proc.	limited	not set
Italy	No	1991 IStat 1% sample ²	In proc.	limited	not set
UK	No	1991 SAR ²	In proc.	limited	not set

Notes to Table 18-3:

- ¹ Indicates whether a sample corresponds to the final version of PAU's recommended sampling strategy. 'Yes' signifies that it does. 'Partially' signifies that the sample has been drawn in accordance with an earlier version of the PAU's recommendations, i.e. it covers only the population over age 50 and the persons residing with them (Estonia, Latvia, and Lithuania have provided the entire population over age 50, while Finland has sampled it with progressive oversampling). 'No' indicates that the sample provided has not been drawn specially for this project, and most often covers the entire population without any oversampling.
- ² Identifies the general purpose sample submitted as part of a country's participation in the project.
- ³ Identifies the conditions governing the release of a country's microdata sample. 'General' signifies that a country has signed the project's standard data release arrangement. So far nine countries have signed the standard arrangement. Hungary and Switzerland (marked as 'limited') require a clearance to be obtained from their national statistical offices for the use of micro data, however the documents signed between the PAU and these countries include clauses stipulating that, in general, all scholars interested in social research will be granted access. Russia (also marked as 'limited') requested that certain provisions for archiving the micro data samples be removed from its data release arrangement. The conditions under which the National Statistical Office of the United Kingdom will release microdata to foreign researchers are still under negotiation; however the PAU has an agreement with several British scholars to facilitate access to the 1991 SARS through collaborative arrangements.
- ⁴ Although the data processing for Latvia and Lithuania is considered to be completed, there are still certain outstanding issues that are being clarified with the respective national statistical offices.

Confidentiality Provisions

The arrangements governing the release of data assembled as part of this project stipulate that they should be used solely for social and behavioral science research. The data and other materials are also to be used solely for statistical

analysis and reporting of aggregated information, and not for investigation of specific individuals or organizations.

Therefore, the user promises to not use these data for purposes other than social and behavioral science research. The user must understand also that attempts to disseminate copies of the data, in whole or part, or to identify individual respondents may be subject to prosecution under applicable laws. Additionally, the user of these data must agree to send a copy of any publications based on the data to the Population Activities Unit.

Research Possibilities

The datasets are suitable for analyzing a wide range of topics within the social sciences, although their true *forte* is that they capture better than most other data sets the social, demographic and economic aspects of population ageing. Being census based these samples do not include the substantive depth and detail of specialized surveys. They have, however, several important advantages compared to other data sources: (1) because of the high sampling density they cover various small population groups, information on whom is not commonly available in other sources; (2) these samples contain as much geographic detail as possible under each country's confidentiality requirements; (3) they include more extensive information on housing conditions than many other data sources; and (4) they provide information for a number of countries whose data were not accessible until recently. Some of the research possibilities offered by these data are illustrated by the existing publications (see bibliography).

Expert Users

Susan De Vos (University of Wisconsin, USA) is a expert on modern international population and housing censuses, and has worked with the samples from PAU's collection available through ICPSR/NACDA.

Kalev Katus and Allan Puur (EKDK, Estonia), experts on Soviet population statistics and censuses, assisted with the

recording of the four samples in PAU's collection that are based on the 1989 census of the former Soviet Union.

Data Expansion

The dataset will be expanded to include the 2000 round of censuses of population and housing as they become available. Meanwhile, processing and standardization of 1990 censuses continues for those countries that remain incomplete.

Bibliography

Although the dataset is still under development, they have already resulted in several publications.

- De Vos, Susan and Gary Sandefur. 1999. "Elderly Living Arrangements Bulgaria, the Czech Republic, Estonia, Finland, and Romania." Madison: CDE Working Paper 99-14, University of Wisconsin.
- Gierveld, J., H. de Valk, and M. Blommesteijn. 2000. "Living Arrangements of Older Persons and Family Support in More Developed Countries." Paper presented at United Nations Technical Meeting on Population Ageing and Living Arrangements of Older Persons: Critical Issues and Policy Responses (New York, 8-10 February 2000) [UN/POP/AGE/2000/5]
- Legare, Jacques, et al. 1998. *Living Arrangements of Older Persons in Canada: Effects on their Socio-economic Conditions*. New York and Geneva: United Nations.
- Lindgren, J., A. Miettinen and M. Nieminen. 1999. *Socio-economic Status and Living Arrangements of Older Persons in Finland*. New York and Geneva: United Nations.

