Elzbieta Golata
Jan Paradysz
Department of Statistics
The Poznan University of Economics
Poland

Indirect Estimation in Population Census 2011 in Poland

The aim of this paper is to discuss the need, conditions and possibilities of introducing new census methodology in Poland. According to preliminary methodological guidelines issued by Central Statistical Office [*Narodowy Spis Powszechny* ..., 2007], the next population and housing census in Poland is planned to be conducted under the register-based censuses design with sample surveys. The virtual census using available registers will construct a micro-data base upon which the census tables will be produced. And the register-based census will be complemented by sample surveys. To provide estimates in detail geographical division indirect estimation techniques will be applied.

Indirect Estimation in Demography

The need to provide population counts for detailed geographic division of a country as well as for domains defined by age, sex, marital status and other demographic variables, in the between census years, led to the development of special indirect estimation methods. They are based on exploration of administrative files that contain valuable demographic information related to population changes and construction of explicit linking models. Among others, the symptomatic accounting techniques [Bogue, 1950; Bogue and Duncan, 1959; Purcell and Kish, 1979], regression symptomatic procedures [Schmitt and Crosetti, 1954; O'Hare, 1976; Ericksen, 1974] and dual-system estimation of population [Rao, 2003] could be mentioned.

But application of indirect estimation techniques in demography is much wider. Conducting a demographic analysis it is natural to refer to the development of a corresponding processes in other populations which were characterized by a similar demographic situation in past. It is also common to analyze data from different, available sources like population census, vital statistics, statistical registration or specific surveys. A review of specific demographic indirect techniques is provided in UN manual [UN, 2005]. The indirect estimation in demography is applied in the absence of vital statistics and the lack of population structures. An example may be estimation of fertility based on information on children ever born according to E. Arriaga [UNPF, 1993], L. Henry [1953] method for estimating fertility within marriage and own children method developed by G. Feeney [1975], W. H. Grabill, Lee-Jay Cho and D.J. Bogue [Grabill, Cho 1965; Grabill, Cho, Bogue, 1970].

The essence of indirect estimation is data comparison, comparison of survey methodology applied, analysis of the definitions used in different surveys, seeking for the differences, their causes, and taking all of these into account, implementation of such estimation techniques that cut down the gap between available statistical data and the required information. The traditional small area estimation methods in demographic research employ indirect estimators based on models utilizing symptomatic variables which are strongly related to changes in local population. The main data source for auxiliary variables in small area estimation is the population census. The expansion of sophisticated small area estimation methodology results also in the development and implementation of alternative methods for conducting censuses [see UN 2007; *Conference for European Statisticians ...,* 2006]. Typically indirect methods in demography used only administrative and census data, while

the current proposition involves also sample survey data in conjunction with other auxiliary population information. Currently, demographers propose to use the indirect estimation methodology not only on the basis of population census. The novelty is the inverse of heretofore experiences, the new population census methodology based on population registers and indirect estimation based on sampling survey data to provide census tables.

Population Census 2011 in Poland

The method proposed for population census 2011 in Poland is register – based census with sample surveys and application of indirect estimation to provide information for small areas. The long form will be replaced by a large sample survey. Implementation of a short form census with en electronic form, hand held devices and internet return of forms is also considered.

Detail analysis and evaluation of the previous, 2002 Population Census is regarded as the starting point for preparing a register based Population Census 2011. The 2002 Census was conducted by using traditional methodology and the administration registers were not used. As a result significant differences between numbers of population given by Census and Population Register (PESEL) were observed, especially at different levels of territorial division, in example in large cities. Experimentally self-enumeration was introduced in 2002 Census. But it was negatively evaluated due to inadequately prepared self-completion questionnaire. The results of post-enumeration survey were not published. The above reasons motivated work on methods ensuring better accuracy of the forthcoming census estimates.

An extensive work is carried by the Central Statistical Office to capture all census core topics and as many as possible non-core topics, from different available administration registers, integrate the electronic databases, coordinate the cooperation of the database administrators and check for the consistency of the information from different sources. The preliminary virtual census based on administrative records is constructed and carefully examined. Problems arising while estimating accuracy of data from the virtual census questionnaire constructed upon different administration records concern many topics: quality of administrative records, their content, ensuring conformity of administrative data with the census definitions, their adequacy with UN population census recommendations, techniques and methods for combining data from different sources etc. A guide in this field is provided by the implementation of Virtual Census in 2001 in Statistics Netherlands [*The Dutch Virtual Census ...*, 2004] and the experience of Statistics Finland - the world pioneer in the statistical use of administrative data sources [*Use of Register and Administrative ...,* 2004].

One of the utmost importance problems refers to the census enumeration date. As administration records are updated at different time, an important task is to estimate their data, especially population structures, for the census enumeration date. From the demographic point of view concerning age and sex distribution, the minimum requirement would be semi-annual update. For that reason special multi-regional life tables should be constructed and implemented. Particular attention is required for the elder age groups and infant mortality. It includes models and methods of disaggregating coefficients for five or even one year intervals into days with respect to seasonality effects.

Advantages and disadvantages of using administrative records for taking a census are discussed. Comparison analysis and a deep discussion evaluating different data sources including omissions, false enumerations, dynamic changes of their accessibility and quality as well as other properties should be provided.

Indirect Estimation in Population Census 2011

Introducing register – based census involves also application of a large sample survey providing long form census statistics. This involves application of indirect estimation

methodology for effective population estimates for small areas. Auxiliary information will be derived from different administration registers and additional data sources.

The integration of statistical data from several sources towards an integrated 'synthetic' data base is a well-known problem in the compilation of statistics. F. Willekens [1994] mention three examples: the compilation of an input-output table for the national accounts, the weighing of micro-data to make a sample representative and the construction of synthetic estimators for the combination of data from a register and a household survey. Data from the different sources are usually contradictory to each other and the evaluation problems arise. As one of the solutions used to combine data from different sources it is proposed to construct an estimator being a function of data from both sources [Huigen et al., 1988, p.28]. The principles and requirements concerning the 'synthetic' micro-data base for census estimates which will combine data from administration registers, short-form census and long-form survey are being carefully formulated. Integrating these data from may arise also weighting problems.

Constructing models for small area estimates and examination of their stability in time and in territorial division will be provided on the basis of the integrated micro-data base. Different approaches are considered: individual and overall evaluation of the models. Among questions formulated in this topic are: What is more important, not the best but good evaluation of precision measures for all areas, or choosing the best models for selected areas of special importance. Subsequently delimitation of similar regions for which different models and estimators might be applied is contemplated. Similar problems refer to stability in time. Additionally some model experiments refer to urban audit data, which may provide census estimates for cities involved in this European project.

Standard indirect estimation techniques like synthetic and empirical bayes are experimentally implemented for different types of small domains. Testing of small area estimation techniques is based on data from the previous 2002 census, surveys conducted within census – like female fertility survey as well as national surveys like household budget survey HBS and labour force survey LFS which were conducted during the census year. The study comprise overall evaluation of the results obtained from domain specific point of view and combining all domains.

Poland belongs to countries which are developing, and testing alternative techniques for collecting, processing and disseminating statistics generated by the population and housing censuses. Before new methods are implemented an overall evaluation is needed to meet the crucial demand - to provide reliable and precise population characteristics at the lowest level of geographical desegregation.

Bibliography:

Bogue D.J., 1950, *A Technique for Making Extensive Postcensal Estimates,* "Journal of the American Statistical Association", 45, s. 149 – 163.

Bogue D.J., Duncan B.D., 1959, *A Composite Method of Estimating Post Censal Population of Small Areas, Sex and Colour,* Vital Statistics-Special Report 47, No. 6, National Office of Vital Statistics, Washington DC.

Conference for European Statisticians Recommendations for the 2010 Censuses of Population and Housing, 2006, Geneva, http://www.unece.org/stats/publ.htm

Ericksen E.P., 1974, *A Regression Method for Estimating Population Changes of Local Areas,* "Journal of the American Statistical Association", 69, s. 867 – 875.

Feeney G., 1975, *The Own-Children Method of Estimating age Specific Fertility Rates: Lecture Notes Prepared for A Demographic Analysis and Data Evaluation Workshop,* Department of Statistics, Kuala Lumpur, Malaysia, 10-28 March, 1975, G. Feeney, The East-West Center, Honolulu, Hawaii, 96822,.

Grabill W.H., Cho L.-J., 1965, *Methodology for the measurement of current fertility from population data on young children*, "Demography", No. 1, s. 50-73.

Grabill W.H., Cho L.J., Bogue D.J., 1970, *Differential Current Fertility in the United States,* Community and Family Study Center, University of Chicago Press, Chicago.

Henry L., 1953, *Fécondité des mariages. Nouvelle méthode de mesure*, "Travaux et Documents", No. 16., Presses Universitaires de France, Paris.

Huigen, R. D., van de Stadt, H. and Zeelenberg, C., 1988. `Socio-economic accounts'. Netherlands Official Statistics 3(2), p. 38 .

Narodowy Spis Powszechny Ludnosci i Mieszkan 2011. Założenia metodyczne, 2007, Material na posiedzenie Rady Programowej narodowego spisu powszechnego ludnosci i mieszkan 2011 r. 4 września 2007, GUS, Warszawa.

O'Hare W., 1976, *Report on Multiple Regression Method for Making Population Estimates,* "Demography", 13, s. 369-379.

Purcell N.J., Kish L., 1980, *Postcensal estimates for local areas (domains),* "International Statistical Review", Vol. 48, s. 3-18.

Rao J.N.K., 2003, *Small Area Estimation*, Wiley.

Schmitt R.C., Crosetti A.H., 1954, *Accuracy of the Ratio Correlation Method of Estimating Poscensal Population,* "Land Economics", 30, s. 279-280.

UN, 2005, *Manual X: Indirect Techniques for Demographic Estimation. Population and Development in the United Nations System*, http://www.un.org/esa/population/publications/Manual_X/Manual_X.html

UN, 2007, *United Nations Principles and Recommendations for Population and Housing Censuses,* 2007, UNPF, 1993, *Readings in Population Research Methodology,* published for United Nations Population Fund by Social Development Center, Chicago, Illinois.

Use of Register and Administrative Data Sources for Statistical Purposes. Best Practices of Statistics Finland, 2004, ed. Pekka Myrskylä, Statistics Finland.

Willekens F., 1994, *Monitoring International Migration Flows in Europe. Towards a Statistical Data Base Combining Data From Different Sources*, "European Journal of Population", Vol. 10, No 1a, s. 1-42.

The Dutch Virtual Census of 2001. Analysis and Methodology, 2004, eds. Eric Schulte Nordholt, Marijke Hartgers, Rita Gircour, Statistics Netherlands.