STATISTICS IN TRANSITION-new series, December 2010
Vol. 11, No. 3, pp. 639—647

BOOK REVIEW

National Research Council (2009). *Coverage Measurement in the 2010 Census.* Panel on Correlation Bias and Coverage Measurement in the 2010 Decennial Census, Robert M. Bell and Michael L. Cohen (Eds.). Committee on National Statistics, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press, 180 pages

Methods of data quality assessment of population censuses were discussed at the Committee of Demographic Research of the Polish Academy of Sciences during last three years (Dygaszewicz, 2008; Kordos, 2008, 2010; Nowak, 2008), and presented in several Polish statistical publications (Gołata, 2008, 2009; Kordos, 2007; Paradysz, 2008, 2010). It is mainly connected with preparation for the 2011 Census of Population and Housing in Poland, and increasing concern in quality of statistical data.

Preparation for the US 2010 Census of Population started directly after the 2000 Census (National Research Council, 2000). Among other problems to be solved, the coverage measurement in the 2010 Census has been considered as the most important. The National Research Council appointed a Panel on Coverage Evaluation and Correlation Bias. Findings of the National Research Council's Panel are published in the book "*Coverage Measurement in the 2010 Census"* with Robert. Bell and Michael L. Cohen, as Editors.

This book consists of five chapters, three appendixes, and nearly 100 references and bibliography, including 13 recommendations for the Census Bureau, which is in charge for US Censuses of Population. This book, which details the findings of the National Research Council's Panel on Coverage Evaluation and Correlation Bias, strongly supports the Census Bureau's change of goal. However, the panel finds that the current plans for data collection, data analysis, and data products are still too oriented towards measurement of net coverage error to fully exploit this new focus. Although the Census Bureau has taken several important steps to revise data collection and analysis procedures and data products, this book recommends further steps to enhance the value of coverage measurement for the improvement of future census processes.

Chapter 1

Introduction. This introductory chapter stressed that since the 1950 census there has been efforts by the Census Bureau to estimate the size of error in census

counts for areas and demographic groups and to use the information to improve census processes. The programs to measure census coverage error are referred to as coverage measurement programs. In recent years, coverage measurement programs included a third objective—correcting the census for enumeration error, referred to as census adjustment. The techniques used in coverage measurement programs to understand the extent of enumeration errors are sample surveys, dual-systems estimation (DSE), and demographic analysis.

This introductory chapter is followed by four chapters and three appendices:

Chapter 2

Fundamentals of Coverage Measurement. Coverage measurement is a collection of techniques that measure the differences between census enumerations and the corresponding true counts for groups or areas. Coverage measurement is the quantitative aspect of coverage evaluation, which also encompasses more qualitative techniques, such as ethnographic observation. The differences between census counts and the corresponding true counts at the level of the individual (or the household) are referred to collectively as census coverage errors, and in this chapter are categorized as types of census coverage error and indicate methods that can be used for their summarization. They then detail the three primary (potential) uses of census coverage measurement that rely on summarizations. Finally, they provide a brief overview of the methods that are currently used in the U.S. census for coverage measurement. This chapter also presents short histories of the U.S. census coverage measurement programs from 1950 to 1990, including a description of accuracy and coverage evaluation program, the coverage measurement program for the 2000 census.

Chapter 3

Plans for the 2010 Census. This chapter examines how the 2010 census differs from the 2000 census with respect to the impact on the coverage measurement program for 2010. It looks in some depth at the treatment of duplicates in the 2010 census and the 2010 coverage measurement program, including the possibility of contamination of the 2010 coverage measurement data collection through the application of the coverage follow-up interview. The chapter also discusses how the use of administrative records could potentially assist in both coverage improvement and coverage measurement for the 2010 census.

Chapter 4

Technical Issues. Here, a number of technical topics introduced by various changes made in coverage measurement for 2010 is discussed, including: (i) the sample design for the census coverage measurement post enumeration survey in 2010; (ii) the use of logistic regression modeling as a substitute for post stratification in modeling net coverage error; (iii) how one compares competing models in this situation; and (iv) the treatment of missing data in net coverage

error modeling, including the Census Bureau's current plans for addressing missing data prior to fitting the logistic regression models in 2010. In relation to the issue of missing data, the chapter includes a description of an attempt by the Census Bureau to greatly reduce the number of cases that are considered to have insufficient information to support matching. The chapter concludes with a discussion of how to improve demographic analysis for use in census coverage measurement in 2010.

Chapter 5

Analytic Use of Coverage Measurement Data. First, the Census Bureau's framework for defining and estimating components of census coverage error is briefly outlined. Then, potential variables for use in statistical models to assess correlates of components of census coverage error are considered. The chapter ends with a consideration of the purpose of the key output from the census coverage measurement program in 2010—the analytic capability to develop statistical models linking census coverage errors of various types to individual and household characteristics and census process variables.

There are three appendixes:

Appendix A

A Framework for Components of Census Coverage Error.

A major goal and challenge for coverage measurement in 2010 is to design a survey that measures the components of coverage error, namely erroneous enumerations and omissions. The Census Bureau's previous coverage measurement surveys were designed primarily to estimate net census error using Dual System Estimation (DSE). To improve the accuracy of estimates of net error, the Census Bureau's DSE has relied on balancing some of the components of error, meaning some census omissions offset some erroneous inclusions in a manner that preserved the net error. As a result, the process produced inflated estimates of omissions and erroneous inclusions. This appendix summarizes Mulry and Kostanich (2006) paper, which provides a framework for overcoming these inflated estimates of component errors. It also explicitly defines the individual components of error and how these components relate to traditional net error concepts.

Appendix B

Logistic Regression for Modeling Match and Correct Enumeration Rates.

This is very technical appendix which provides details on the use of logistic regression models as a substitute for post stratification. More information is available in: (Malec, D., and Maples, 2005; Mulry et al., 2005; Schindler, 2006). This research suggests that inclusion of small-area effects could substantially improve coverage estimates. Several questions remain: how best to treat the complex sample design, how many random effects can be included and at what level of aggregation, the best way to estimate the model parameters, and how the

model fit should be assessed. The panel is impressed with this high-caliber research that addresses an important issue in coverage modeling; further work in this area would be very valuable.

Appendix C

Biographical Sketches of Panel Members and Staff, provides biographical sketches of panel members and staff.

The panel offers 13 recommendations concerning coverage measurement plans for 2010.

To achieve this new goal, instead of only measuring net census error, the Census Bureau also plans to measure the four components of census coverage error: (1) census omissions, (2) census duplications, (3) erroneous census enumerations, and (4) census enumerations in the wrong location. The panel supports these plans, since different types of coverage errors are caused by different interactions between census processes and housing units and their occupants. The estimation of these four components of coverage error can be supported by the general structure of the data collection and matching that is carried out in support of dual-systems estimation, though modified and expanded to support this different purpose. The panel finds, however, that the Bureau's plans could be more fully developed for this purpose. Additionally, the panel recommends to allocate sufficient sources for research program on decennial census improvement in future.

Recommendation 1

The Census Bureau should more completely shift its focus in coverage measurement from that of collecting data and developing statistical models with the goal of estimating net coverage error to that of collecting data and developing statistical models that support the improvement of census processes.

Recommendation 2

The Census Bureau should allocate sufficient resources, including funding and staff, to assemble and support an ongoing intercensal research program on decennial census improvement. Such a group should focus on using the data from the census and the census coverage measurement programs to identify deficient census processes and to propose better alternatives. The work of this group should be used to help design the census tests early in the next decade.

Recommendation 3

The Census Bureau should retain comprehensive data on the functioning of the coverage follow-up interviews for a substantial sample of cases, especially for those cases in the CCM block clusters, to support detailed follow-up analysis of the functioning of the follow-up interviews and to help suggest modifications and alternatives for use in 2020.

Recommendation 4

The Census Bureau should organize census and coverage follow-up data collection so that data collection for the census coverage measurement (CCM) program is initiated as soon as possible after the completion of the census. In particular, the post enumeration survey in a particular area should start as soon as possible after the completion of the great majority of the census data collection—hopefully before late July. The Census Bureau should also consider census designs for 2010 in which there is some modest overlap between coverage follow-up and CCM data collections.

Recommendation 5

The Census Bureau should use the various testing opportunities in both the 2010 census and in the early part of the 2010–2020 intercensal period to assess how administrative records can be used in the 2020 census.

Recommendation 6

The Census Bureau should compare its sample design for the 2010 census coverage measurement post enumeration survey with alternative designs that give greater sampling probability to housing units that are anticipated to be hard to enumerate. If an alternative design proves preferable for the joint goals of estimating component coverage error and net coverage error estimation, such a design should be used in place of the current sample design.

Recommendation 7

The Census Bureau should develop missing data techniques, in collaboration with external experts if needed, which preserve associations between imputed and observed variables, condition on variables that are predictive of the missing values, and incorporate imputation uncertainty into estimates of standard errors. These ideas should be utilized in modeling the census coverage measurement data collected in the 2010 census.

Recommendation 8

The Census Bureau should give priority to research on improving demographic analysis in the four areas: (1) improving the measurement of undocumented and documented immigrants, (2) development of sub-national geographic estimates, (3) assessment of the uncertainty of estimates from demographic analysis, and (4) refining methods for combining estimates from demographic analysis and post enumeration survey data.

Recommendation 9

The Census Bureau should further develop and refine its framework for defining the four basic types of census coverage error and measuring their frequency of occurrence. The Census Bureau should also develop plans for

operationalizing the measurement of these components using data from the census and the census coverage measurement program.

Recommendation 10

In developing the logistic regression models or other types of discriminant-analysis models of match status, correct enumeration status, and components of census coverage error, the Census Bureau should consider:

- Use of several approaches before focusing on a specific model; besides logistic regression, alternatives should include use of other link functions, discriminant analysis, and various data mining approaches, such as classification trees, support vector machines, and neural nets.
- Thorough examination of the subset of predictors that is best suited to each individual statistical model; the predictors for these various statistical models need not be identical; however, there may be a benefit to constraining the (logistic regression) models of match rate and correct enumeration rate to have identical variables in the estimation of net coverage error, and research should be carried out to assess whether this benefit outweighs the benefit of selecting variables that are optimal for each of these two logistic regression models.
- To effectively blend information from auxiliary sources at various levels of geographic and demographic aggregation, random effects modeling and Bayes' methods also should be examined.

Recommendation 11

The primary output of the Census Bureau's coverage measurement program in 2010 should be an analytic database that is used to support the development of statistical models to inform census process improvement. The production of summary tabulations should be of lesser priority.

Recommendation 12

The Census Bureau should develop regression models that elucidate the various types of census coverage error, using specified dependent and predictor variables. To the extent that the database supporting these models can be made available to external researchers, it is extremely important that the Census Bureau pursue all viable avenues to involve outside researchers in the development of such models.

Recommendation 13

For a sample of households, the Census Bureau should retain data that provide a comprehensive picture of the census processes used to enumerate it, and the individuals residing in it, to facilitate subsequent evaluation. To allow linking assessment of census coverage error with a history of the census processes, this sample should substantially overlap with the CCM sample.

Improvement the quality of subsequent censuses

Although it is important to assess census coverage, it would also be extremely helpful to use that assessment to improve the quality of subsequent censuses. Consequently, an important use of coverage measurement is to help to identify important sources of census coverage errors and possibly to suggest alternative processes to reduce the frequency of those errors in the future. Although drawing a link between census coverage errors and deficient census processes is a challenging task, the Census Bureau thinks that substantial progress can be made in this direction. Therefore, the 2010 coverage measurement program has the goal of identifying the sources of frequent coverage error in the census counts. This information can then be used to ***allocate resources toward developing alternative census designs and processes*** that ***will provide counts with higher quality in 2020***. It is conceivable that use of such a feedback loop could also provide substantial savings in census costs, in addition to improvement in census quality because the trade-off between the effect on accuracy and on census process costs might now be better understood. The panel fully supports this modification of the objectives of coverage measurement in 2010.

For our readers we would like also to recommend a new published monograph by UN Statistics Division (2010) on ***Post Enumeration Survey***, which is strictly connected with the reviewed book.

Prepared by Jan Kordos, Warsaw School of Economics,
E-mail: jan1kor2@aster.pl

REFERENCES

DYGASZEWICZ, J. (2008), *Project of the 2011 Census of Population and Housing in Poland – the Most Important Strategic Aims*. The Committee of the Demographic Research of the Polish Academy of Sciences (in Polish). http://www.knd.pan.pl/images/stories/pliki/pdf/Dygaszewicz_luty_2008.pdf

GOŁATA, E. (2009), Economic activity in population census 2011 and administration resources. In: Gołata, E. (Ed), *Methods and Sources of Obtaining Information in Public Statistics,* Wydawnictwo Uniwersytetu Ekonomicznego w Poznaniu. Poznań. (in Polish).

GOŁATA, E. (2010), Indirect Estimation of Economic Activity for the Register-based Census. In: Gołata, E.(Ed), *Measurement and Information in Economy*, Wydawnictwo Uniwersytetu Ekonomicznego w Poznaniu, Poznań, pp. 85–104 (in Polish).

HOGAN, H. (2003). The Accuracy and Coverage Evaluation: Theory and Design. Survey Methodology, 29(2):129-138.

KORDOS, J. (2007), Some Aspects of Post-Enumeration Surveys in Poland, *Statistics in Transition-new series,* December 2007, Vol. 8, No. 3, pp. 563–576.

KORDOS, J. (2008), Methods of Census Data Quality Assessment – Current Practice in Poland, and Suggestions Connected with the 2011 Population Census Preparation (in Polish). http://www.knd.pan.pl/images/stories/pliki/pdf/

KORDOS, J. (2010), Methods of Quality Assessment of Population Census Data. http://www.knd.pan.pl/images/stories/Census-Q-_sl_-2011.pdf

MALEC, D., and MAPLES, J. (2005). An evaluation of synthetic, small-area census coverage error using a random effects model. Proceedings of the Section on Survey Research Methods of the American Statistical Association. Available: http://www.amstat.org/Sections/Srms/Proceedings/y2005/Files/JSM2005-000180.pdf.

MULE, T. (2007). Coverage measurement for the 2010 census. Presented at the Washington Statistical Society Meeting, Washington, DC, January 23, 2008. Available: http://scs.gmu.edu/~wss/wss080123slides.pdf [accessed 11/25/08].

MULRY, M.H., SCHINDLER, E., MULE, T., NGUYEN, N., SPENCER, B.D. (2005). Investigation of extreme estimates of census coverage error for small areas. Proceedings of the Section on Survey Research Methods of the American Statistical Association: http://www.amstat.org/Sections/Srms/Proceedings/y2005/Files/JSM2005-000551.pdf.

MULRY, M.H., and KOSTANICH, D.K. (2006). Framework for census coverage error components. Proceedings of the Section on Survey Research Methods of the American Statistical Association. Available: http://www.amstat.org/sections/srms/Proceedings/y2006f.html

NATIONAL RESEARCH COUNCIL. (2000). Redesigning the 2010 Census: First Interim Report. Panel on Research on Future Census Methods, Committee on National Statistics. Washington, DC: National Academy Press.

NOWAK, L. (1998), Quality of Census Data, In: Tendencies of Changes in Structure of Population, Households and Families in 1998-1995 GUS, Warsaw, pp. 22–31 (in Polish).

NOWAK, L. (2008) , Th e 2011 *Census of Population and Housing – Methodology and Topics*, the Committee of Demographic Research of the Polish Academy of Sciences (in Polish). http://www.knd.pan.pl/images/stories/pliki/pdf/Nowak_28_luty_2008.pdf

PARADYSZ, J. (1989), On non-sampling errors in women's fertility survey in the 1970 National Population Census. In: *Problems of Statistical Surveys by Sampling*. GUS, Warszawa, BWS, vol. 36, pp. 154–159. (in Polish).

PARADYSZ, J. (2010), Necessity of Indirect Estimation in National Census. In: Gołata, E. (Ed); *Measurement and Information in Economy*, Wydawnictwo Uniwersytetu Ekonomicznego w Poznaniu. Poznań, pp. 45–66 (in Polish).

SCHINDLER, E. (2006). 2010 census coverage measurement: The hunt for the magic variables. Proceedings of the Section on Survey Research Methods of the American Statistical Association. Available: http://www.amstat.org/sections/srms/Proceedings/y2006f.html.

UN STATISTICS DIVISION (2010), *Post Enumeration Surveys*, Operational guidelines, Technical Report, New York, April 2010.