Integrating Disability Census Microdata: What is accessible from IPUMS-International?

Robert McCaa and Krishna Mohan Palipudi, University of Minnesota Population Center

1. Disability. Empirical studies on disability across countries and over time may provide useful information for designing focused interventions and policy recommendations. To better understand health outcomes as well as inequalities, it is essential to measure these using reliable data sources (Wang, 2002). One of the major concerns in carrying out cross country analysis is the reliability as well as the comparability of data sources, both across countries and over time. This problem is particularly acute for disability estimates.

The census is used by many countries to collect data on persons with disabilities and for some countries such data have been collected through the census for a long time. Although limited in terms of accuracy and coverage, for many countries the census still provides the only source of information on number of persons with disabilities and their social and economic characteristics.

The paper offers a brief overview of the IPUMS-International census microdata integration project and discusses the availability and use of comparable census microdata on disability across countries and over time. The paper focuses on two related issues. The first is to provide information and discussion on IPUMS data dissemination and to offer some conclusions and ideas on the methodology and policy issues related to integration of disability micro data in IPUMS international and secondly, to analyze data on disability prevalence and identify some of the methodological issues that need to be addressed in order to improve the quality and comparability of census data on disability.

2. The IPUMS-International collaboratory. The IPUMS-International initiative is a global collaboratory of universities, National Statistical Offices, and international research institutes to preserve, integrate and manage access to high-density census microdata samples (Ruggles et. al. 2003). Begun in 1999 with funding provided by the National Institutes of Health and the National Science Foundation of the United States, to date the project enjoys the endorsement of National Statistical Offices (NSOs) in sixty-seven countries, encompassing more than sixty percent of the world's population. Fifty-eight NSOs have entrusted microdata to the project for a total of 172 censuses. Currently, integrated samples are accessible from www.ipums.org/international for 20 countries, 63 censuses and 185 million person records. Densities for most of the samples are ten percent, although some are five percent and a few are even less.

Documentation is essential to the success of the initiative. The IPUMS-International project archives, scans and posts census documentation on the internet. Currently, at www.hist.umn.edu/~rmccaa/IPUMSI/enumform.htm, enumeration forms for 749 censuses may be viewed and downloaded. Instructions to enumerators are available for several hundred censuses as well. Before census microdata are made accessible for any census, unstinting efforts are made to obtain copies of complete documentation. These are posted on the project website, in the official language and English translation, at the same time that the microdata are released to users. The IPUMS Dynamic Metadata System facilitates the comparison, in English, of any census question in the database for any combination of

countries and census years. By means of a few simple clicks the researcher may compare not only the wording on the enumeration form but also the instructions to the enumerators for any desired combination of censuses in the database.

3. IPUMS Integration Methods. To make census microdata useful for research they must be not only thoroughly documented, but also integrated. IPUMS-International adopts uniform coding schemes, nomenclatures and classifications, based where possible on the United Nations Statistics Division's *Principles and Recommendations for Population and Housing Censuses* (first 1998, now 2006) and other international standards. Censuses employ differing nomenclatures and numeric classification systems and reconciliation of these codes is a major effort. Variables must be easy to use for comparisons across time and space. This requires that we provide the lowest common denominator of detail that is fully comparable. On the other hand, we must retain all meaningful detail in each sample, even when it is unique to a single dataset (Esteve and Sobek 2003, McCaa et. al. 2006).

In the case of disability variables, we settled upon a yes/no integration classification scheme for each type: disabled, blind, deaf, mute, lower extremities, upper extremities, mental, psychological, personal care, mobility, public transportation, work, etc. Then to retain all detail, researchers are offered the original nomenclatures in the "unharmonized" variables. The former may be useful for comparison, while the later is most useful for analyzing a single census, yet may be used by the researcher to develop a unique integration scheme for two or more censuses. In either case, the researcher is obligated to compare the documentation to understand the nuances of how questions were posed and whether variables are comparable between one census and another. Due to the great variety of phrasings, notwithstanding considerable international effort at standardization, researchers are urged to use the "unharmonized variables" when studying disability in the IPUMS system.

4. Comparability between countries (2000 census round only): The IPUMS international has used post-harmonization technique (Rijckevorsel, 2001) to transform some extent incomparable data into a comparable version. The big advantage of it is that we can use the existing data, but we do not know whether applying such transformations will affect the results. It is understandable that with out making stronger and explicit assumptions related to concepts and design of questions, it is impossible to establish a micro data that has cross-cultural applicability. So, it is high time to examine such data sets and provide proper documentation to users in order to be useful for policy oriented research.

One of the major problems in comparing the disability statistics from different censuses is the lack of consistency and terminological uses and classification (Bartley, 2001). There is a variety of information in our collection on how disability was defined in national censuses. What is available shows that disability is defined differently in different countries. For example, with regard to 2001 round census, in South Africa, disability is defined as a 'serious condition that prevented the respondent's full participation in life activities such as education, work and social life'. In Ecuador the term disability is defined as a 'permanent difficulty in doing an activity that is considered normal, due to irreversible effects from an incurable congenital or acquired disease' In Philippines 2001, it is defined as 'any restriction or lack of ability (resulting from an impairment) to perform an activity in the manner or within the range considered normal for a human being'. Uganda speaks to a long term physical condition or health problem

lasting 6 months or more. The problem is further compounded by the unsuitability of reference period and questions used in assessing disability across countries prior to 2000 census rounds. As can be seen from these examples, disability in Ecuador was defined in terms of activity limitations, in Philippines in terms of impairments and in South Africa in terms of participation. The definitions used in national data collection activities may be influenced by the use of the data as well as cultural practices and perceptions in the countries concerned.

With regard to the questions used to identify the population with disabilities, there are differences in: (a) the type of questions used, i.e., whether impairment, activity limitations or participation based; (b) the wording of the questions with regard to terms used; (c) the scope of the questions in terms of the number of disability items included and (d) the reference period that was considered to determine a persons disability status (Mathiowetz, 2001). According to the available literature, the questions asked to identify the population with disabilities in censuses were broadly categorised into 4 types (Mbogoni and Angela Me, 2002).

Type 1 - A generic/general question on presence of a condition combined with items on participation and activity limitations.

Type 2 - A generic/general question on presence of the disabled or handicapped in the household followed by a list of impairments and/or disabilities.

Type 3 - A checklist of impairments from which respondents are required to choose.

Type 4: Employment or work related questions used to assess the disability.

Examples of questions related to the four types and a list of countries that fall in to these categories are reported in the main paper. As these questions suggest, there is a lack of uniformity in the questions used by countries to identify the population with disabilities. There are substantial differences between the censuses in the rates of disability prevalence partly as a result of methodological differences with respect to how disability is defined; the design of questions used to identify the population with disabilities; and the type of disabilities included. Countries that have used type 3 questions have higher prevalence rates than the other types, for example South Africa, 2001 (5.1%); Ecuador, 2001 (4.68%); and Uganda, 2001 (3.47%) compared to the type 2 questions used in Philippines, 2000 (1.53%); Uganda, 1991 (1.14) and type 4 from Venezuela, 1981 (1.39).

5. Comparability between censuses within a single country: To assess the extent to which the countries have comparable time series samples and issues involved in such comparisons, we have reviewed some of the questions on disability used in countries that had their census during the 2000 round of censuses. It was observed from the census samples that for certain countries the definitions used, number questions and wording of questions on disability has changed over time. For example, in Uganda it was asked that 'is there anyone who was in the household on census night disabled and the nature of disability' where as in 2002 sample it was restructured as 'does (name) have any difficulty in moving, seeing, hearing, speaking or learning, which has lasted or is expected to last 6 months or more'. In these samples we have observed a significant difference in prevalence of disability. In countries like Chile the content and wording was same for the samples in 1992 (2.14%) and 2002 (2.2%), where the comparisons are more

fruitful. In Philippines, we have samples from three consecutive censuses. All the three samples have administered 2 questions each on obtaining the information on disability. So, we can easily compare the time series data in Philippines at least for the prevalence of over all disability.

Further, we examined differentials in disability estimates across various socioeconomic and demographic subgroups over time in most comparable census samples. The paper also explains the changing inequalities in disability estimates using concentration index in order to show how comparable are the time series samples in IPUMS international data.

6. Conclusions: (just before challenges for the 2010 round) IPUMS data are a useful and important source on disability that provides information on frequency and distribution of disability in the population across countries and over time at national and regional/province level. The study suggests that the prevailing scheme of harmonization, though the estimates vary by question type and definitions used, is useful in comparing time series data to get a better picture on variations in the prevalence of disability as well as using the time series data across countries for the monitoring and evaluation of services concerning the equalization of opportunities. Whatever concepts, nomenclatures, and coding schemes are used in the census operations for the 2010 round of censuses, IPUMS will retain the originals nomenclatures in the "unharmonized" variables. For the integrated variables an attempt may be made to go beyond the yes/no classification to a more detailed composite coding scheme. Official statisticians and researchers are invited to use the IPUMS data and documentation and make suggestions to enhance not only the treatment of disability variables but also the IPUMS system as a whole

References:

Esteve, Albert and Matthew Sobek. (2003). Challenges and Methods of Census Harmonization. *Historical Methods* 36: 66-79.

Mathiowetz N.A. (2001). Methodological issues in the measurement of persons with disabilities. *Research in Social Science and Disability*, Vol. 2, 125-143.

Mbogoni Margaret and Angela Me, (2002). Revising the United Nations Census Recommendations on Disability. Paper prepared for the First meeting of the *Washington Group on Disability Statistics*, Washington, 18-20 February 2002.

McCaa, Robert, Albert Esteve, Steven Ruggles and Matt Sobek. (2006) Using integrated census microdata for evidence-based policy making: the IPUMS-International global initiative. *African Statistical Journal* 2: 83-100

Monica Bartley (2001). Pros and Cons of Data Collected Using a Census: The experience of the Caribbean Countries. *International Seminar on the Measurement of Disability*, New York.

Rijckevorsel van J.L.A. (2001). Cross population comparison of surveys: A review of new technologies. *International Seminar on the Measurement of Disability*, New York.

Ruggles, Steven, Miriam King, Deborah Levison, Robert McCaa, and Matthew Sobek. (2003). "IPUMS-International: An Overview". *Historical Methods*, 36: 60-65

Wang Liming (2003). Determinants of child mortality in low-income countries; Empirical findings from Demographic and Health Surveys. *Health Policy*. Sep; 65(3):277-99.