Journal of Demographic Economics, **81**, 2015, 203–216.
doi:10.1017/dem.2014.6

THE IPUMS COLLABORATION: INTEGRATING AND DISSEMINATING THE WORLD'S POPULATION MICRODATA

Steven Ruggles, Robert McCaa, Matthew Sobek and Lara Cleveland
University of Minnesota, Minneapolis, Minnesota, USA

Abstract: The Integrated Public Use Microdata Series (IPUMS)-International partnership is a project of the Minnesota Population Center and national statistical agencies, dedicated to collecting and distributing census data from around the world. IPUMS is currently disseminating data on over a half-billion persons enumerated in more than 250 census samples from 79 countries. The data series includes information on a broad range of population characteristics, including fertility, nuptiality, life-course transitions, migration, labor-force participation, occupational structure, education, ethnicity, and household composition. This paper describes sample characteristics and data structure; the data integration process including the creation of constructed family interrelationship variables; the flexible dissemination system that enables researchers to build customized extracts of pooled census samples across time and place; and some of the most significant findings that have emerged from the database.

Keywords: census microdata, data harmonization, data dissemination, demographic analysis

1. INTRODUCTION

Two decades ago, census microdata were rare. In most countries, census statistics were available only in aggregated tables that described the characteristics of places. In the United States, however, census microdata had been a mainstay of social science research ever since 1962, when the U.S. Census Bureau released a one-in-thousand sample of the household and person records that had been collected in the 1960 census enumeration. The impact on American social science was profound; as sociologist Otis Dudley Duncan expressed it, "the importance of this innovation can hardly be overestimated. ... all too often efforts to put information into an appropriate form are frustrated by the inadequacy of the published summary tables for the purpose at hand. With access to the unit records, the social scientist may specify in detail how variables are to be manipulated so as to produce an optimal estimate" (Duncan 1974: 5097).

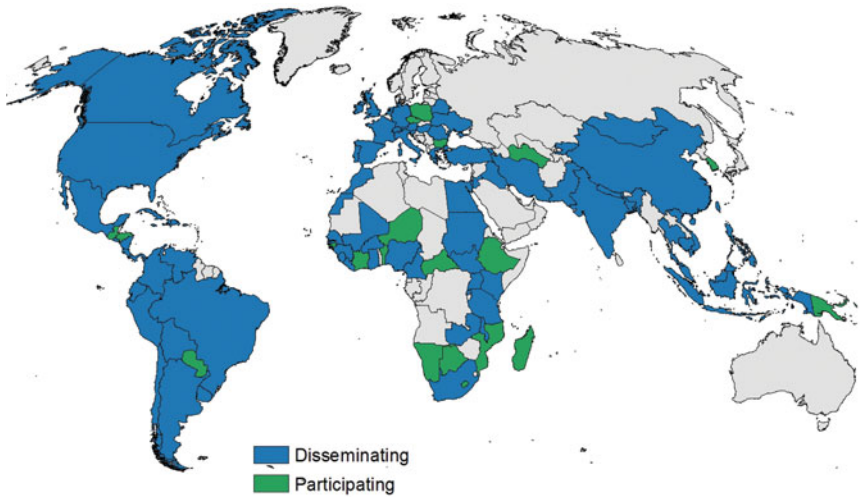

FIGURE 1. (Colour online) IPUMS participating countries.

Outside of the United States, statistical agencies were slow to make microdata available to the research community. Canada began producing limited microdata files in 1974 and the United Kingdom followed suit in 1993, but in both cases the samples were small and had limited variables, and the statistical agencies restricted access to within-country researchers. In a few other countries, large samples of high-quality microdata were available to selected researchers through special arrangements with statistical agencies, but international comparative research remained difficult or impossible. Moreover, most countries had no systematic program for preservation or re-use of census microdata once the statistical agency had published summary tables. As a result, much machine-readable census microdata from the 1960s and 1970s had already disappeared by the mid-1990s. Much of the surviving census microdata were at risk of loss through deterioration of the storage media or retirement of technical staff needed to locate and interpret the files (McCaa 2013; Ruggles 2014).

The IPUMS project began in 1992, and was initially focused on U.S. microdata. IPUMS released an integrated set of microdata from ten U.S. censuses in 1995, and it was popular from the outset: by making the entire series of censuses easily interoperable, IPUMS saved researchers redundant effort. In 1999, we received funding from the National Science Foundation and the National Institutes of Health to add microdata from other countries, beginning with Colombia, France, Kenya, Mexico, and Vietnam. Over the next 15 years, we formed partnerships with 100 national statistical agencies around the world. Figure 1 shows the geographic distribution of IPUMS partners. The countries labeled "disseminating" are those with data available through IPUMS as of September 2014; the additional countries labeled "participating" have joined the collaboration, but their data and metadata are still being processed. The project depends on the active collabo-

https://doi.org/10.1017/dem.2014.6 Published online by Cambridge University Press

ration of both the statistical agencies and international statistical organizations, who not only contribute data, but also provide invaluable expertise needed to process them. A complete list of national and international partners is available at https://international.ipums.org/international/international_partners.shtml.

We are currently disseminating data on 560 million persons from 258 samples of 79 countries. Variables are coded consistently across censuses, enabling researchers to readily make comparisons between countries and across time periods. A web-based data access system allows users to easily browse this vast database, selecting only those records and variables necessary for their analysis. Researchers may either download customized subsets for local analysis or analyze the data without downloading them, by using IPUMS online data analysis tools. Researchers must apply for access, demonstrating a reasonable scientific need for the data; but once approved, they have access to the entire database. The data access system is available at www.ipums.org/international.

2. GOALS

IPUMS seeks to preserve, integrate, and disseminate census and survey microdata to qualified researchers without cost, while ensuring the confidentiality of respondents. The most urgent goal is to preserve the microdata. Much data have been lost over the years, and more is at risk of destruction. National censuses are collected at great public expense, but they often fall into neglect once the results are published and a national statistical office turns its attention to the next census. The data are of little use without the metadata explaining how to interpret them; both must be preserved if future generations are to benefit from these rich data sources. IPUMS carefully archives all data and metadata entrusted to us by statistical agencies, ensuring their long-run survival. In many cases, IPUMS has funded the recovery of old data off of aging magnetic tapes that were otherwise unreadable by their statistical offices.

Integration of data and metadata is the core activity of the IPUMS project. We harmonize the data so the same codes mean the same things for all times and countries in the database. We confidentialize the data to remove the possibility of identifying individuals; this is critical for public trust and is stipulated by the dissemination agreements with the statistical offices that provide the data. Finally, the IPUMS collaboration is committed to the goal of democratizing access to data. All IPUMS data are available to researchers everywhere free of cost, to anyone with an Internet connection and a viable research project. No one is privileged with special access, and no one conducting legitimate research is turned away.

3. DATA

Each IPUMS record represents an individual and is composed of variables describing that person's characteristics as collected by the relevant census. With a few exceptions, individuals are organized into households, and within households we

https://doi.org/10.1017/dem.2014.6 Published online by Cambridge University Press

TABLE 1. Number of IPUMS samples by country (258 Total)

Argentina	5	Fiji	5	Malawi	3	Senegal	2
Armenia	1	France	7	Malaysia	4	Sierra Leone	1
Austria	4	Germany	4	Mali	3	Slovenia	1
Bangladesh	3	Ghana	2	Mexico	7	South Africa	3
Belarus	1	Greece	4	Mongolia	2	South Sudan	1
Bolivia	3	Guinea	2	Morocco	3	Spain	3
Brazil	6	Haiti	3	Nepal	1	Sudan	1
Burkina Faso	3	Hungary	4	Netherlands	3	Switzerland	4
Cambodia	2	India	5	Nicaragua	3	Tanzania	2
Cameroon	3	Indonesia	9	Nigeria	5	Thailand	4
Canada	4	Iran	1	Pakistan	3	Turkey	3
Chile	5	Iraq	1	Palestine	2	Uganda	2
China	2	Ireland	9	Panama	6	Ukraine	1
Colombia	5	Israel	3	Peru	2	UK	2
Costa Rica	4	Italy	1	Philippines	3	USA	7
Cuba	1	Jamaica	3	Portugal	3	Uruguay	6
Dominican Rep.	5	Jordan	1	Puerto Rico	5	Venezuela	4
Ecuador	6	Kenya	5	Romania	3	Vietnam	3
Egypt	2	Kyrgyzstan	2	Rwanda	2	Zambia	3
El Salvador	2	Liberia	2	Saint Lucia	2		

can identify family interrelationships. This data structure provides substantially more power than would a simple sample of individuals. Thus, a researcher has access not only to the person's characteristics, but to all the characteristics of the people with whom they lived, family and non-family. This allows the construction of new variables drawing from information across individual person records, such as the number of wage earners in a family, or whether a mother has children under age five. Each set of persons also has a household record that contains information shared by household members, such as geography, and—in most censuses—the attributes of the dwelling in which they lived, such as presence of piped water or number of rooms.

IPUMS currently includes 258 microdata samples from 79 countries, as shown in Table 1. Many countries not included in Table 1 have already sent data to the project that are currently awaiting processing. IPUMS processes on average 25 new samples every year. In all cases the samples are nationally representative. The modal sample density is 10% of the national population, but 5% is also common, and some samples are lower density. The median sample size is 828,000 person records, and the database in total has 560 million persons. Roughly two thirds of the samples are from developing countries. The temporal scope of the database is currently from 1960 to the present, but we are adding new samples for earlier census years. Because most countries have samples from multiple censuses, it is usually possible to analyze change over time nationally and internationally. It is not possible to track individuals across censuses.

https://doi.org/10.1017/dem.2014.6 Published online by Cambridge University Press

TABLE 2. Selected topical coverage of harmonized IPUMS variables

Household characteristics	Person characteristics
Geography	**Migration**
First-administrative level	Previous residence
Second-administrative level	Years in current locality
Urban-rural status	**Fertility/mortality**
Dwelling	Children ever born
Number of rooms	Children surviving
Toilet access	Parental mortality
Construction materials	**Nativity/ethnicity**
Age of structure	Place of birth
Living area	Country of birth
Utilities	Citizenship
Electricity	Year of immigration
Water	Religion
Sewage	Race
Fuel	Ethnic group
Heating	Language spoken
Amenities	Mother tongue
Automobiles	**Education**
Washer	School attendance
Television	Literacy
Computer	Educational attainment
Phone	Years of schooling
Other	**Labor force**
Home or land ownership	Employment status
Number of deaths	Occupation
Number of international migrants	Industry
Family and household composition	Class of worker
	Hours worked
Person characteristics	Total income
Core demographic	Wage and salary income
Age	Source of livelihood
Sex	**Disability**
Marital status	Disability status
Age at marriage	Type of disability
Relationship to householder	Cause of disability

Table 2 describes the topical scope of IPUMS samples. The datasets generally include information on economic activities, ethnicity, educational attainment, fertility, migration and place of former residence, marital status and consensual unions. Many developing countries provide information about mortality and disabilities, and there are extensive housing characteristics, usually including water supply, sewage, and physical characteristics of the dwelling such as floor and roof

https://doi.org/10.1017/dem.2014.6 Published online by Cambridge University Press

materials and number of rooms. The census-by-census availability of particular variables can be viewed at https://international.ipums.org/international/.

4. DATA PROCESSING

IPUMS processing has three major components: reformatting and correction of format errors; metadata integration; and variable integration. The following paragraphs briefly describe each of these processes.

4.1. Reformatting and Correction of Format Errors

Systematic reformatting and cleaning of each dataset involves analyzing the record structure, reformatting the data into a standard hierarchical format, applying internal consistency checks, and correcting data errors. The oldest datasets—from the 1960s and 1970s—pose the greatest problems, a consequence of the computing constraints of the time. Even the most recent samples, however, require effort to verify that they are free of data format problems.

The microdata come in a variety of formats (e.g., rectangular, hierarchical, linked files). We reformat each sample into a simple hierarchical structure consisting of a household record followed by person records for each individual in the household. Any geographic or dwelling-level information is replicated on each household record. This reformatting often exposes problems that cannot be identified from a detailed examination of data frequencies and is an integral aspect of diagnosis and cleaning. Some statistical agencies draw samples for IPUMS, but often they provide us with entire censuses. In these cases, we draw 10% geographically-stratified samples.

Most of the data files that serve as the raw material for IPUMS—even those with the highest data quality—have never been cleaned to meet the standards necessary for public-use datasets. We have built procedures for detecting and correcting common data errors into our processes for adding samples to the IPUMS database. For example, we check for such things as households with no householder or multiple householders, householders with multiple spouses in countries without polygamy, implausibly large households, and duplicate records. When feasible, we correct the records based on logical inference using other information in the household.

4.2. Metadata Integration

Data are useful only when researchers understand what they mean. Accordingly, we have developed comprehensive harmonized documentation on each variable and sample. This documentation covers enumeration procedures and instructions; definitions of households, dwellings, group quarters, and other enumeration units; and scanned images of original-language versions of the questionnaires. We also provide detailed descriptions of each variable, including question wording and

https://doi.org/10.1017/dem.2014.6 Published online by Cambridge University Press

instructions (in the original and translated into English), universe definitions, frequency distributions, and variable codes. Comparability discussions describe any deviations of particular censuses from the standard variable definition and address differences over time and across countries.

The scale of the documentation is substantial. For the present collection of censuses and surveys, we have prepared approximately 250,000 words documenting characteristics of the samples and a million words describing variables. We have also translated and tagged three million words of census and survey forms and instructions to enumerators. If published in conventional printed form, this documentation would require over 10,000 large-format pages. The facsimiles we provide of the original-language forms and instructions for each instrument would add about 7,000 additional pages of material. To make the large scale of documentation usable, all information about IPUMS datasets is converted into structured metadata that can be processed by machine (Sobek, Hindman and Ruggles 2007).

4.3. Variable Integration

All variables in the international census samples are numerically coded, and the classifications are inconsistent across census years and countries. Reconciling these codes is a major part of the project. We retain all the details provided in the original samples, except where confidentiality edits are needed. At the same time, we provide a truly integrated database, in which identical categories in different samples receive identical codes. We employ several strategies to achieve these competing goals of maximizing comparability and retaining detail. For simple variables, such as age and sex, the original variables are compatible, and recoding them into a common classification is straightforward. For more complicated variables, it is impossible to construct a single uniform classification without losing information. Some censuses provide more detail than others, so the lowest common denominator of all samples inevitably loses important information. In these cases, we construct composite coding schemes. The first one or two digits of the code provide information available across all samples. The next one or two digits provide additional information available in a broad subset of samples. Finally, trailing digits provide detail only rarely available (Esteve and Sobek 2003; Ruggles 2006).

The classification scheme for marital status illustrates the approach. Under the IPUMS design, the first digit of marital status has four categories consistently available in all samples: (1) single, (2) married/in union, (3) separated/divorced/spouse absent, and (4) widowed. The distinction between divorced and separated is not maintained in all samples, so these categories are combined in the fully comparable first digit. At the second digit, we distinguish divorced and separated persons in the samples with that information, as well as formal marriages and consensual unions. The third and final digit differentiates among types of marriages (civil, religious, polygamous) available for selected countries only.

https://doi.org/10.1017/dem.2014.6 Published online by Cambridge University Press

For each variable, we develop integration metadata that provide information on the location of the original variable in each sample, each original category value, category labels in the original language and in English, and each new standardized category value and label. The system can accommodate complex recodes in which information from more than one variable in the original census is needed to construct a new compatible variable.

5. DISSEMINATION

The IPUMS data are accessed via a web-based dissemination system. This is the only way the data are distributed—no one has early access or can obtain data not available to every other researcher. The data access system allows users to design datasets that are customized to their particular research problem, by merging data across time periods and countries, selecting population subsets, selecting variables, and defining new variables that capitalize on the hierarchical structure of the data. Users design their dataset in a rich informational environment that describes each sample and variable, with special attention to the comparability of particular items across time and space. Although researchers can conduct online data analysis, most download and analyze the data on their own computers using the software package with which they are most familiar. The IPUMS system supports SAS, SPSS, and STATA, and we plan to add R in the near future.

IPUMS is designed to facilitate cross-national research. The data extract system lets users define pooled datasets that include any variables they desire from as many times and places as they wish. Thus, country and year can be variables in the analysis. Using the extract system it is feasible to build a single dataset containing selected variables for all 560 million persons in the database. If such a dataset would be too large, the system is capable of drawing a systematic subsample of cases. Of course, most analyses are more localized in time and place, but IPUMS offers the unique potential for truly globe-spanning research. This is a practical possibility not only because of the data extract system, but also due to the harmonization of the variable codes and to the documentation system that collates information at the variable level across samples. Thus, the primary logistical barriers to cross-national studies are removed, and researchers can focus on the substantive matters of interest to them.

The hierarchical structure of IPUMS makes it possible to interrelate the characteristics of co-resident persons in creative ways. To fully exploit this feature of the data, IPUMS constructs "pointer" variables that identify the location within the household of each person's mother, father, and spouse, if they were present. This makes it simple to compare the characteristics of spouses, to attach parents' characteristics to children or vice versa, and to construct unique household or family-level measures. For example, one can make a variable for spouse's education, mother's birthplace, or father's migration status. The IPUMS data access system allows users to construct such variables automatically.

https://doi.org/10.1017/dem.2014.6 Published online by Cambridge University Press

IPUMS has prepared GIS boundary files that enable users to map the data for all countries at the first administrative level (states, provinces, etc). The first-level has also been harmonized across time within countries, so users can be certain that the same geographic codes describe the same space in all periods. The project is currently engaged in a longer-term goal of providing boundaries and harmonized geography for the second-level (e.g. counties) within as many countries as possible. The geographic work will be leveraged by a major related project at the Minnesota Population Center: Terra Populus (https://www.terrapop.org/). Terra Populus allows users to add environmental data from satellites and climate models to the person records in IPUMS. Thus, a researcher can make a variable for the percent of forest cover or annual rainfall in a geographic unit, and have that appear as a variable on each person's record.

In another major new dissemination initiative, IPUMS is developing a restricted system that will give researchers access to high-density and even full-count data with full geographic detail. Access sites are strictly controlled, and the data can be analyzed only remotely on the IPUMS servers. No data will be transmitted—only the results—and those will be subject to review by IPUMS staff before their release to ensure confidentiality protection. The restricted access system will allow new kinds of analyses that use small places to study such things as human-environment interactions, health outcomes related to location, segregation, or access to services.

6. IPUMS USERS AND USES

Over 10,000 researchers have registered to use the international IPUMS data, and they have produced about 1,000 articles and working papers. Economists comprise the largest disciplinary group of users, accounting for nearly 40% of the total. The IPUMS has stimulated exciting and creative new research on economic development, population growth and movement, fertility, mortality, nuptiality, and family demography, as well as the economic and social correlates of demographic behavior and the causes and consequences of demographic change. The data are especially valuable for studying trends and differentials in the core demographic processes of fertility, mortality, migration, marriage, and family composition, and have become a major source for the reports of the U.N. Population Division (Gerland *et al.* 2013). The paragraphs that follow outline the strengths and limitations of the microdata in each of these areas.

6.1. Fertility and Mortality

IPUMS offers multiple strategies for studying fertility and mortality. The most widely-used approach for fertility analysis is the own-child method, which uses the age differentials of mothers and children to estimate age-specific fertility estimates (Cho, Retherford and Choe 1986). IPUMS provides a variety of variables to simplify own-child fertility analysis. For children, IPUMS identifies the mother's record number. For mothers, IPUMS provides the number of children under five

https://doi.org/10.1017/dem.2014.6 Published online by Cambridge University Press

in the household, age of eldest child, and age of youngest child. Most IPUMS samples also include information on children-ever-born, which allows analysis of cohort parity distributions (David *et al*. 1988). Many of the samples also include information on births to each woman within the past year, allowing direct estimation of fertility rates.

The most important approach for IPUMS mortality analysis capitalizes on the widespread availability of information on the number of children surviving for each mother; in combination with the information on children-ever-born, these data can be used to generate robust estimates of age-specific mortality of children (United Nations 1983; Preston and Haines 1991; Hill 2013). Following United Nations recommendations (United Nations 2008) an increasing number of developing countries have added questions on deaths of household members in the preceding 12 months. Although this source understates overall mortality levels—partly because persons living alone are not counted—the data have nevertheless proven to be an invaluable source (Dorrington, Moultrie and Timæus 2004; Garenne, McCaa, and Nacro 2008). Finally, some countries have added questions about maternal or paternal orphanhood, which can also be used for indirect mortality estimation (United Nations 1983).

The alternatives to IPUMS for fertility and mortality analysis are vital statistics and demographic surveys. Compared with vital statistics, the advantage of IPUMS is that it allows individual-level analysis of covariates. For example, one may simultaneously assess differential fertility or child mortality by education, husband's occupation, ethnicity, and household composition. The chief disadvantage of IPUMS compared with vital statistics is that vital statistics may give more reliable estimates of fertility levels. This disadvantage mainly applies to developed countries that have high-quality registrations systems; in the developing world, vital registration systems are often substantially worse than are the results from the census.

Demographic surveys often have more comprehensive fertility histories than are available in the census, and they usually have questions about contraception and fertility intentions that are not ordinarily available in censuses. The greatest advantage of IPUMS compared with fertility surveys is sample size, which permits analysis of small population subgroups at single years of age. Moreover, IPUMS allows fine-grained spatial analysis of fertility spanning multiple decades; such analysis is typically impossible with survey data. Surveys also have limited use for analyzing mortality, since mortality analysis requires much larger samples.

6.2. Migration

IPUMS is well suited to the study of migration, with most censuses including one or more questions on the topic. The most widely available migration variables are of two general types: place of birth and place of residence at some time prior to the census. Each type records internal as well as international migration. Birthplace data identify lifetime migrants without accounting for return migration

https://doi.org/10.1017/dem.2014.6 Published online by Cambridge University Press

or intervening moves. Period migration variables typically record residence one or five years prior to the census, with some censuses reporting a person's previous residence irrespective of when they left it. Additional, less frequently available variables record such attributes as nationality, year of immigration, and reason for migration. Some censuses enable tracking migration steps and return migration by using multiple variables.

There are few alternatives to census data for studying migrant populations. Flow measures can be calculated from effective population registers or immigration records, but only the census provides adequate cases to measure migrant stocks and, more importantly, the characteristics of specific migrant populations. Common topics of study include migrant attainment (Spörlein and van Tubergen 2014), endogamy (Choi and Mare 2012), brain drain (Docquier and Marfouk 2006), schooling (Halpern-Manners 2011), and gender effects (Donato 2010). Because IPUMS has data from both sending and receiving countries, it is possible to study the characteristics of migrants compared to those they left behind (Feliciano 2005). The data also allow the study of internal migration on a multi-national scale (Bernard, Bell and Charles-Edwards 2014).

6.3. Marriage, Cohabiting Unions, and Family Composition

IPUMS is the most powerful source available for comparative analysis of changing patterns of marriage, cohabiting unions, and family composition. All IPUMS samples have information on marital status, and almost all identify consensual unions. In addition, many samples have information on age at first marriage, date of first marriage, or marriage duration. Because of the shift from marriage to cohabiting unions around the world [e.g. Esteve, Lesthaeghe and López-Gay (2012)], statistics derived from marriage certificates are no longer useful for the study of union formation.

IPUMS provides flexible tools for studying family composition. Because individuals are listed with all their co-resident family and household members, and within families the relationship of each individual to the householder or household head is known, researchers can construct virtually any measure of family and household composition. The IPUMS family interrelationship variables—spouse's location in household, mother's location in household, father's location in household, and family unit identifier—make it simple for researchers to construct customized measures of living arrangements tailored to particular research questions.

IPUMS provides a cross-sectional view: longitudinal surveys are needed to study marriage and family transitions. Such surveys, however, are available for only a limited set of countries. Moreover, they often are not comparable with one another, and they are poorly suited to the study of period change. For consistent analysis of marriage and family behavior around the world over decades of dramatic change, there is no real alternative.

https://doi.org/10.1017/dem.2014.6 Published online by Cambridge University Press

6.4. Exemplary Studies

IPUMS data are shifting the landscape of scientific research on the human population by opening new opportunities for comparative dynamic analysis. It is no longer sufficient to study the relationships among variables in a particular place at a particular moment. To understand the large-scale processes that are transforming the planet, we must investigate processes of change. For most countries, IPUMS is the only available source of microdata for the study of long-run change, and the IPUMS design makes such investigations comparatively simple.

Consider these award-winning examples:

- Esteve, Lesthaeghe and López-Gay (2012) documented an extraordinary rise of unmarried cohabitation across 350 Latin American regions in 13 countries between 1970 and 2000. Their analysis of individual-level and regional characteristics suggests that the rise of cohabitation was not an expansion of traditional practices; instead, the authors argue, this Latin American cohabitation boom represents a distinctive new phenomenon.
- Lam and Marteleto (2008) used data from eight countries in Latin America, Africa, and Asia to develop a new characterization of the demographic transition from the perspective of children competing for resources within families and cohorts. Although the transition has played out with striking regularity in country after country, the path of changes in cohort size and number of siblings is not linear, owing to the complex interaction of population momentum with falling fertility and mortality.
- Bleakley (2010) compared Brazil, Colombia, Mexico, and the United States to assess how much childhood exposure to malaria depressed labor productivity. Highly effective malaria treatment campaigns in each country allow spatiotemporal analysis of the impact of the disease. The results were remarkably similar across the four countries; exposure to a malaria eradication program was associated with a 25% increase in earnings as an adult.
- Nawrotzki, Riosmena and Hunter (2013) combined IPUMS data with precipitation data to show that changes in rainfall are inversely associated with U.S.-bound migration. Accordingly, diminishing the vulnerability of rural Mexican households to ecological change through mechanisms such as sustainable irrigation and drought-resistant crops could slow northward migration.

The international IPUMS database has become a vital element of our shared scientific infrastructure because it provides a unique laboratory for the analysis of economic and social processes and offers the empirical foundation we need for developing and testing theoretical models. Microdata are vital for understanding powerful large-scale trends such as economic development, urbanization, fertility transition, migration, population aging, and mass education. These data are also uniquely suited for assessing the *consequences* of social, economic, and demographic transformations in such diverse areas as family structure, economic inequality, and cultural diversity and assimilation. Perhaps most important, cross-temporal and cross-national data on the human population are crucial for understanding changes in the earth's interconnected biological and climate systems.

https://doi.org/10.1017/dem.2014.6 Published online by Cambridge University Press

By creating a framework for locating, analyzing, and visualizing the world's population in time and space, these data provide unprecedented opportunities to investigate the agents of change, assess their implications for human society and the environment, and develop policies to meet future challenges.

REFERENCES

Bernard, A., M. Bell and E. Charles-Edwards (2014) Life-course transitions and the age profile of internal migration. *Population and Development Review* 40, 213–239.

Bleakley, H. (2010) Malaria eradication in the Americas: A retrospective analysis of childhood exposure. *American Economic Journal: Applied Economics* 2, 1–45.

Block, W. and W. L. Thomas (2003) Implementing the data documentation initiative at the minnesota population center. *Historical Methods* 36, 97–101.

Cho, L. J., R. D. Retherford and M. K. Choe (1986) *The Own-Children Method of Fertility Estimation*. Hawaii: East-West Population Institute.

Choi, K. H. and R. D. Mare (2012) International migration and educational assortative mating in Mexico and the United States. *Demography* 49, 449–476.

Cleveland, L., R. McCaa, S. Ruggles and M. Sobek (2012) When excessive perturbation goes wrong and why IPUMS-International relies instead on sampling, suppression, swapping, and other minimally harmful methods to protect privacy of census microdata. In J. Domingo-Ferrer and I. Tinnirello (eds.), *Privacy in Statistical Databases*, pp. 179–187. Berlin and Heidelberg: Springer-Verlag.

David, P. A., T. A. Mroz, W. C. Sanderson, K. W. Wachter and D. R. Weir (1988) Cohort parity analysis: Statistical estimates of the extent of fertility control. *Demography* 25(2), 163–188.

Docquier, F. and A. Marfouk (2006) International migration by education attainment, 1990–2000. In C. Ozden and M. Schiff (eds.), *International Migration, Remittances and the Brain Drain*, pp. 151–199. New York: Palgrave and Macmillan.

Donato, K. M. (2010) U.S. migration from Latin America: gendered patterns and shifts. *Annals of the American Academy of Political and Social Science* 630, 78–92.

Dorrington, R., T. A. Moultrie and I. M. Timæus (2004) Estimation of mortality using the South African Census 2001 data. *Centre for Actuarial Research*, CARe Monograph No. 11, 2004.

Duncan, O. D. (1974) Developing social indicators. *Proceedings of the National Academy of Sciences* 71, 5096–5102.

Esteve, A., R. Lesthaeghe and A. López-Gay (2012) The Latin American cohabitation boom, 1970–2007. *Population and Development Review* 38, 55–81.

Esteve, A. and M. Sobek (2003) Challenges and methods of international census harmonization. *Historical Methods* 36, 66–79.

Feliciano, C. (2005) Educational selectivity in U.S. immigration: How do immigrants compare to those left behind? *Demography* 42, 131–152.

Garenne, M., R. McCaa and K. Nacro (2008) Maternal mortality in South Africa in 2001: From demographic census to epidemiological investigation. *Population Health Metrics* 6, 1–13.

Gerland, P., T. Spoorenberg, J. Bravo, P. Lattes, C. Sawyer, V. Kantorova, and M. S. Lai (2013) Uses of census microdata by the United Nations population division. *XXVII International Population Conference*, Aug 23–25. Busan, Korea.

Halpern-Manners, A. (2011) The effect of family member migration on education and work among nonmigrant youth in Mexico. *Demography* 48, 73–99.

Hill, K. (2013) Indirect estimation of child mortality. In Moultrie, T. A., R. E. Dorrington, A. G. Hill, K. Hill, I. M. Timæus and B. Zaba (eds). *Tools for Demographic Estimation. Paris: International Union for the Scientific Study of Population.* Available at: http://demographicestimation.iussp.org/content/indirect-estimation-child-mortality. Accessed 24/09/2014.

Lam, D. and L. Marteleto (2008) Family size of children and women during the demographic transition. *Population and Development Review* 34, 225–252.

https://doi.org/10.1017/dem.2014.6 Published online by Cambridge University Press

McCaa, R. (2013) The big data revolution: IPUMS-International. trans-border access to decades of census microdata samples for three-fourths of the world and more. *Revista de Demografía Histórica* 30(1), 69–88.

McCaa, R., S. Ruggles and M. Sobek (2010) IPUMS-International statistical disclosure controls. In J. Domingo-Ferrer and E. Magkos (eds.), *Privacy in Statistical Databases*, pp. 74–84. Berlin and Heidelberg: Springer-Verlag.

McCaa, R., S. Ruggles, M. Davern and T. Swenson (2006) IPUMS-International high precision population census microdata samples: balancing the privacy-quality tradeoff by means of restricted access extracts. In J. Domingo-Ferrer and L. Franconi (eds.), *Privacy in Statistical Databases,* pp. 375–382. Berlin and Heidelberg: Springer-Verlag.

McCaa, R. and A. Esteve (2006) IPUMS-Europe: Confidentiality measures for licensing and disseminating restricted access census microdata extracts to academic users. In *Monographs of Official Statistics: Work Session on Statistical Data Confidentiality*, pp. 37–46. Luxembourg: Office for Official Publications of the European Communities. http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.46/2005/wp.5.e.pdf.

Nawrotzki, R. J., F. Riosmena and L. M. Hunter (2013) Do rainfall deficits predict U.S.-bound migration from rural Mexico? Evidence from the Mexican Census. *Population Research and Policy Review* 32(1), 129–158.

Preston, S. H. and M. R. Haines (1991) *Fatal Years: Child Mortality in Late Nineteenth-century America*. Princeton, NJ: Princeton University Press.

Ruggles, S. (2006) The Minnesota Population Center data integration projects: Challenges of harmonizing census microdata across time and place. *Proceedings of the American Statistical Association, Government Statistics Section,* pp. 1405–1415. Alexandria, VA: American Statistical Association. Available at: http://umn.edu/~ruggles/jsmx9.pdf.

Ruggles, S. (2014) Big microdata for population research. *Demography* 51, 287–297.

Sobek, M., M. Hindman and S. Ruggles (2007) Using cyber-resources to build databases for social science research. Minnesota population center working paper series, No. 2007-01. Available at: https://www.pop.umn.edu/sites/www.pop.umn.edu/files/working-papers/Sobek2007-01.pdf.

Spörlein, C. and F. van Tubergen (2014) The occupational status of immigrants in western and non-western societies. *International Journal of Comparative Sociology* 55, 119–143.

United Nations (Population Division) (1983) *Manual X: Indirect Techniques for Demographic Estimation*. New York: United Nations, Department of Economic and Social Affairs, ST/ESA/SER.A/81. Available at: http://www.un.org/esa/population/techcoop/DemEst/manual10/manual10.html.

United Nations (Statistics Division) (2008) *Principles and Recommendations for Population and Housing Censuses*. Statistical Papers Series M No. 67, Revision 2. New York: United Nations, Department of Economic and Social Affairs. Available at: http://unstats.un.org/unsd/publication/seriesM/seriesm_67Rev2e.pdf.

https://doi.org/10.1017/dem.2014.6 Published online by Cambridge University Press