DEVELOPMENT OF AN
INTERNATIONAL POP@RISK GEO-DEMOGRAPHIC DATABASE

Joshua Comenetz
Population Division, U.S. Census Bureau
4600 Silver Hill Road, Washington, DC 20233, USA
Joshua.Comenetz@census.gov

Abstract

The U.S. Census Bureau has a long history of producing annual population estimates by age and sex for every country in the world at the first (province/state) and in some cases second or lower administrative levels. Through the populations at risk (Pop@Risk) effort, the Census Bureau plans to improve access to current international geo-demographic data with the goal of contributing to response to natural disasters and human-induced crises, in Africa and worldwide.

Over the past decade, digital demographic data, boundary maps in GIS formats, and satellite images have become increasingly available. Advances in data accuracy and spatial precision can be achieved by incorporating census data and digital base maps at lower administrative levels, application of improved subnational demographic projection methods, and through use of satellite images to map populated places. Image data can also assist in developing demographic estimates for countries that have no recent census data or limit access to demographic data. This would be of particular value in much of Africa, where population changes and redistribution mean that census results become outdated rapidly. Provided via an internet portal, the Pop@Risk database could be used to improve the efficiency of response to climatic disasters, famines, and other natural or human-induced crises through provision of demographic estimates that are more geographically and temporally accurate than has previously been possible.

Goals

The primary purpose of the international Pop@Risk project is to produce and provide geospatial data and maps that can be used to save lives during and after natural or human-induced disasters. This Census Bureau effort responds to the National Research Council (NRC)'s call for better data and maps in support of disaster response. A recent NRC report included the following recommendations (NRC 2007, 152-55):

> Develop a template of minimum acceptable population and other geospatial data sets that are required by disaster responders. The data sets should be updated frequently (at least mid-decade if not more frequently) and include digital census enumeration units and other census maps in digital form.

Support should be given to test the accuracy of estimates of size and distribution of populations based on remotely sensed imagery, particularly in rural and urban areas of countries with spatially, demographically, and temporally inadequate census data.

The U.S. Census Bureau should be given greater responsibility for understanding populations at risk and should be funded to do so. These responsibilities could include greater capacity and authority for training international demographic professionals in the tools and methods described in this report, and providing data and analytical capabilities to support the U.S. government in international disaster response and humanitarian assistance activities. The U.S. Census Bureau should also have an active research program in using and developing these tools and methods, including remotely sensed imagery and field surveys.

Timely provision of good subnational demographic information can improve the efficiency of humanitarian relief efforts during and after disasters. The Pop@Risk database will contribute to the solution of problems ranging from basic questions—"How many tents should we throw on the plane?"—to longer-term issues such as provision of age-appropriate health care to evacuees. Better targeting of services may improve both the quality of response and allocation of scarce financial resources.

Background

For many years, the Population Division of the U.S. Census Bureau has been a source of international demographic estimates at the national and subnational levels. Annual national estimates and projections (data for entire countries) are included in the International Data Base (IDB) (U.S. Census Bureau 2007). Subnational estimates can be linked to digital maps in ArcGIS shapefile format, allowing for the creation of international demographic maps. Some of these data and maps serve as a base for the LandScan international gridded population database (www.ornl.gov/sci/landscan), created by the Oak Ridge National Laboratory.

Data

The Census Bureau's digital map archive covers every country and territory in the world. Figure 1 shows the smallest administrative unit (ADM) currently held in shapefile format for each country. The goal is to obtain both data and shapefiles at the lowest (smallest) available administrative level (ADM). ADM levels are defined as follows:

- ADM0 – Nations and territories.
- ADM1 – States, provinces, or equivalent.
- ADM2 – Districts, counties, or equivalent.
- ADM3 – Municipalities, townships, or equivalent.
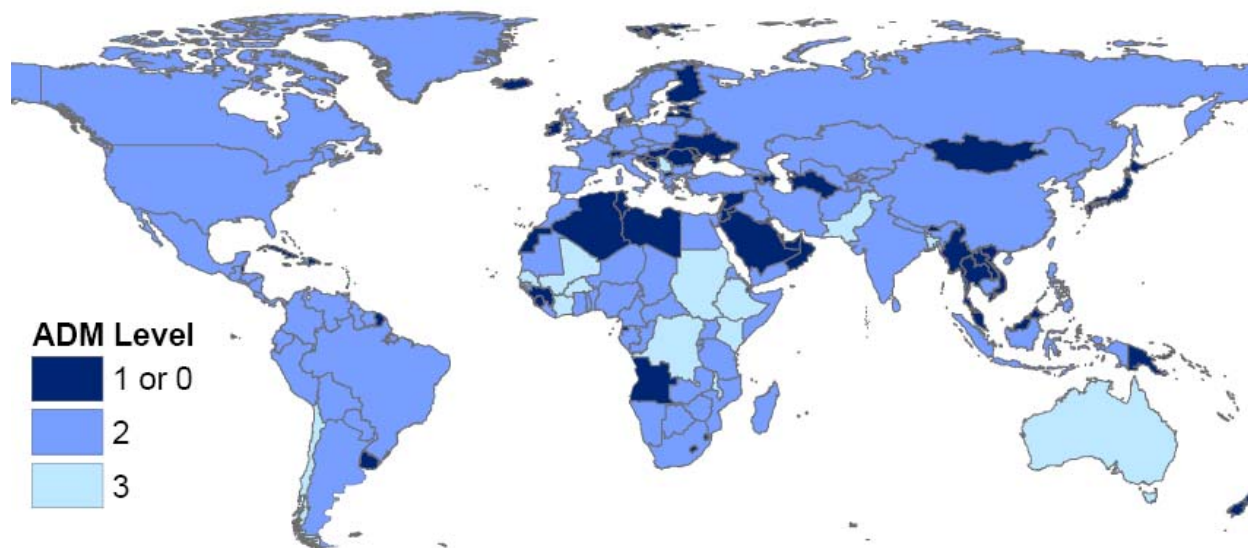- ADM4 and below – Villages or other small units.

Figure 1. Smallest administrative level held in the U.S. Census Bureau digital map archive.

Most of the world is covered at ADM1 and ADM2 levels, but even at ADM2 there are gaps in the Middle East and parts of Asia and Europe. On the other hand, ADM3 maps have been obtained for several countries. Separate maps are obtained for each ADM level wherever possible. While smaller ADMs can usually be aggregated to form higher-level ADM units, this is not always the case, and boundary revisions can occur at any time.   Data sources include the websites of national statistical agencies, international organizations such as the United Nations and World Health Organization, the Department of State and other U.S. government agencies, online collections maintained by academic and other non-governmental organizations, and digitization of paper maps by Census Bureau staff.

The Census Bureau's demographic holdings similarly include population data (total population, and age/sex breakdowns typically in 5- or 10-year bands) for the entire world. In more than 90% of countries or territories, covering over 95% of the world's population, these data are derived from national censuses. Census data are essential to creation and maintenance of a Pop@Risk database because censuses usually cover all of a country's territory, allow for "continuity of statistics from census to census," and provide "detail…about population sub-groups in local areas" (NRC 2002, 7). Where national census data are unavailable or outdated, official estimates or other sources are used—for example, in countries affected by war (Afghanistan, Angola) or where no census has been conducted in more than a decade (Andorra, Greenland).

Figure 2 contains excerpts from the demographic archive. Columns on the left identify geographic areas down to ADM3. The "Geo Match" column provides the unique code used to join population data with shapefile polygons. Total, male, and female population estimates are each broken down by five-year age bands (sample columns shown; full range is 0-4 years, 5-9, 10-14,…75-79, 80 and over).

FIPS	COUNTRY	ADM1 NAME	ADM2 NAME	ADM3 NAME	GEO MATCH	Name Change/ Alternate Name	CONTINENT	IDB YEAR	TOTAL	TOT 0-4	TOT 80+
ET	ETHIOPIA						Africa	2007	76511887	12372169	227988
ET48	ETHIOPIA	DIRE DAWA				DIRE DAWA PROVISIONAL ADM1	Africa		360351	58270	1074
ET48	ETHIOPIA	DIRE DAWA	DIRE DAWA		003ETH015001001		Africa		235858	38139	703
ET48	ETHIOPIA	DIRE DAWA	GURGURA		003ETH015001002		Africa		124493	20131	371
ET49	ETHIOPIA	GAMBELA HIZBOCH				GAMBELLA	Africa		260197	42074	775
ET49	ETHIOPIA	GAMBELA HIZBOCH	ZONE 1				Africa		64679	10459	193
ET49	ETHIOPIA	GAMBELA HIZBOCH	ZONE 1	GAMBELA	003ETH012001004		Africa		37827	6117	113
ET49	ETHIOPIA	GAMBELA HIZBOCH	ZONE 1	ITANG	003ETH012001003		Africa		26852	4342	80
ET49	ETHIOPIA	GAMBELA HIZBOCH	ZONE 2				Africa		51791	8375	154
ET49	ETHIOPIA	GAMBELA HIZBOCH	ZONE 2	ABOBO	003ETH012002005		Africa		19992	3233	60
ET49	ETHIOPIA	GAMBELA HIZBOCH	ZONE 2	GOG	003ETH012002009		Africa		21651	3501	65
ET49	ETHIOPIA	GAMBELA HIZBOCH	ZONE 2	JOR	003ETH012002010		Africa		10148	1641	30

Figure 2. Excerpt from demographic data archive showing 2007 data for parts of Ethiopia.

To provide geo-demographic information that is as accurate and current as possible, both archives—demographic data and shapefiles—require revision and updating as new census results, demographic estimates, and administrative boundary maps are released by national governments. The following complicate this process:

- Demographic data and shapefiles are often produced by different agencies or released separately (on average, it is more difficult to obtain shapefiles than numerical data). If only one of these two elements is available, it is not possible to *map* population. Also, governments may release data and maps at different ADM levels. If one has data at ADM1 only, one cannot map population below ADM1, even if an ADM2 shapefile is available.

- Countries around the world choose to split, combine, or completely redraw administrative areas. For example, Ethiopia (Figure 3) and South Africa redrew provincial and lower-level boundaries in the early 1990s; India and Nigeria periodically create new states; and new counties are occasionally created in the United States.

- Censuses are relatively infrequent—typically a decade or more apart. Most population estimates are based on censuses, and therefore calculated using the administrative boundaries in force at the time that the most recent census was conducted. If boundaries are redrawn after a census, it can be difficult to produce accurate estimates of the populations of new ADMs unless the new ADMs are defined in terms of pre-existing smaller units. For example, if a country redraws ADM1 boundaries but keeps ADM2 boundaries constant, one can calculate the new ADM1 populations; otherwise, this may not be possible.

- Data and maps made available by governments can be withdrawn or revised at any time. Frequent online searches must be conducted, and personal contacts maintained with agencies that produce data, to ensure that all new materials are added to the Census Bureau archive as soon as possible after release or revision.
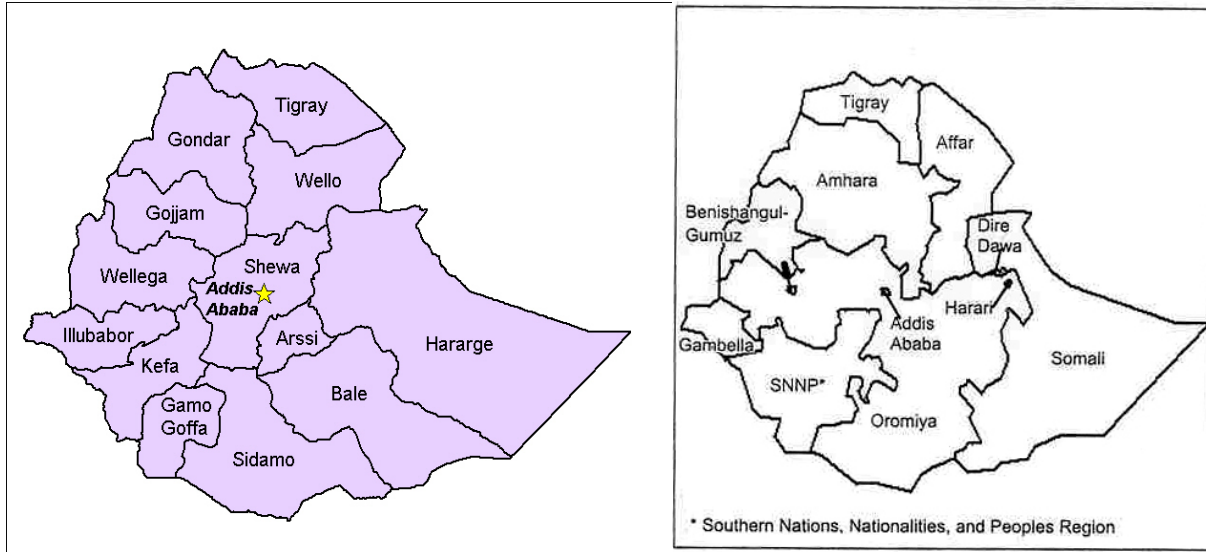

Figure 3. Ethiopian provincial boundaries in effect at time of 1984 census (left) and 1994 census (right). The newer boundaries were drawn along ethnic lines.

The above issues mean that revision and update is a constant process. All available data and shapefiles are collected but matching data and shapefiles frequently are obtained at different times or from different sources. An online Pop@Risk database of demographic maps will therefore reflect the lowest common denominator in both archives—that is, the smallest and most recently delineated administrative units for which both data and digital boundaries are available.

Remote Sensing Methods

The goals of satellite imagery-based population density mapping are to:

- Provide information about the geographic distribution of population where no census has been taken recently, or where census data are no longer reliable. Examples include Somalia, eastern Ethiopia, and Lebanon.

- Map population distribution at finer scales than is possible with census data (Tatem *et al.* 2007). This is necessary where countries do not gather or release data for smaller administrative units.

- Improve the quality of population estimates between census years. Few countries take censuses more frequently than once per decade (UN 2008). Image mapping could help validate demographic projections produced by standard mathematical methods by revealing, for example, rapid growth in urban areas. Image mapping is unlikely to be of as much use in areas of population decline, because the built environment will remain; however, this will not pose a serious problem in the medium term because populations are still growing in most countries.

The methods currently being tested are based on those used to produce the U.S. National Land Cover Dataset for 2001 (Homer *et al.* 2007). First, high-resolution images (e.g., Quickbird, IKONOS, or Orbview, with pixel size of 4 meters or less) for part of the area of interest are obtained (Figure 4, item 1). High-resolution images cover comparatively small areas, and it is difficult or impossible to obtain complete coverage for entire countries. They have the advantage that they show the built environment very clearly, and it is therefore possible to classify each pixel as impervious (hard surface) or pervious (Figure 4, item 2), and to omit non-human impervious surfaces (e.g., bare rock) from the "impervious" category.
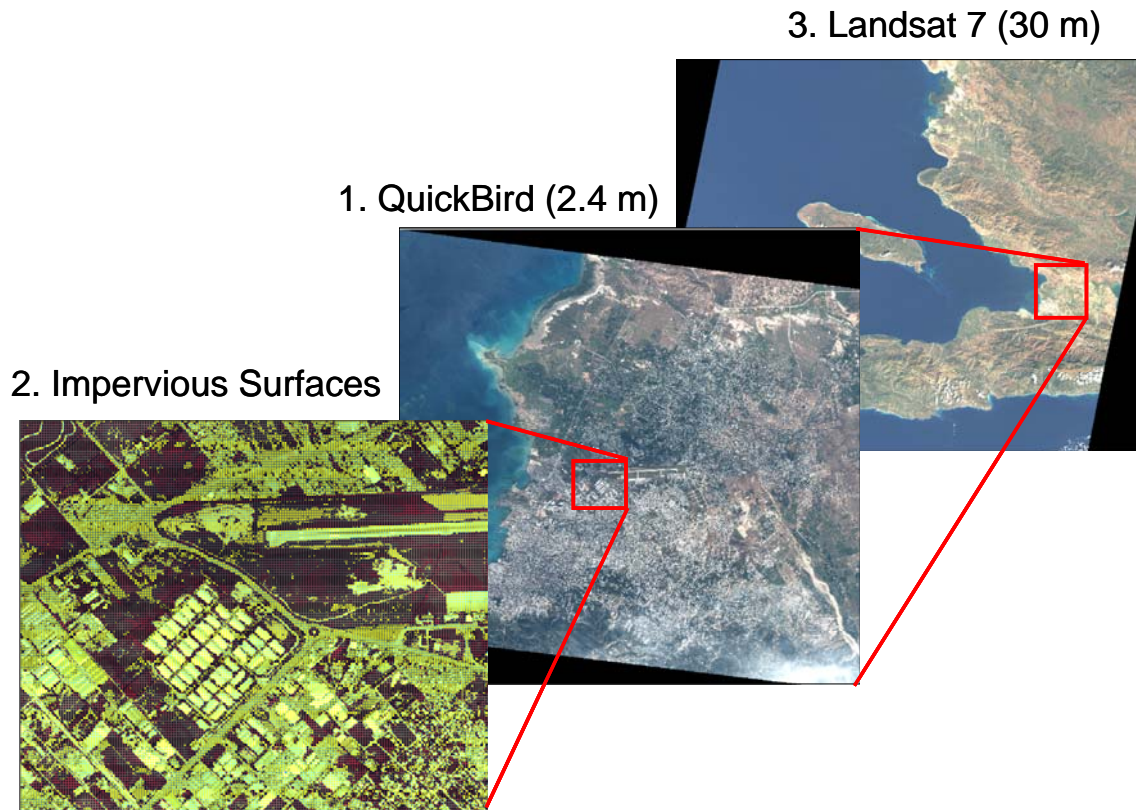

Figure 4. Example of image analysis process, showing high-resolution Quickbird images of Port-au-Prince, Haiti (items 1 and 2), and a lower-resolution Landsat image of the region surrounding Port-au-Prince (item 3).

Next, Landsat images are obtained for the entire country or region of interest (Figure 4, item 3). Landsat images are freely available and have a larger pixel size (30 meters), so relatively few are needed to cover an entire country or disaster-affected region. If it is possible to calculate the percent of impervious surface in each Landsat pixel, then mapping impervious surfaces on a national scale will proceed more rapidly and at lower cost. High-resolution mapping of impervious/pervious surfaces provides training areas: an averaging filter is applied to create 30-meter pixels from the high-resolution results, and the percent of impervious surface is calculated for each of these larger pixels. Next it will be necessary to determine the combination of Landsat bands and elevation/slope characteristics (taken from digital elevation models) that correlates with the percent impervious data derived from the high-resolution images.

The key question is whether the image-based map of percent impervious surface will correlate with population density. Testing the proposed procedure requires:

- Census data, linked to shapefiles for the smallest available ADM units in several countries. The procedure must be tested in regions with different environmental characteristics (desert, humid tropical, temperate, mountainous) and urban planning and building traditions, because varying construction styles are likely to produce different patterns of impervious land cover.

- National-level impervious surface maps of the same countries, derived from satellite images obtained as near to the census dates as possible.

- Information on the location of urban areas in each country. The population density-impervious surface relationship is expected to vary not only among world regions but also within countries, because in poorer countries, for example, rural roads are less likely to be paved than city roads.

Assuming that the above are in hand, the next step is to aggregate Landsat pixels to the smallest available ADM boundaries. It will then be possible to chart the correlation between percent impervious surface and census-derived population density. It may be necessary to separate urban and rural areas to obtain better correlation. The goal is to calibrate image maps using census data, making it possible to estimate population density in areas where images but not census data are available. Among the questions to be answered at this stage are:

- How well do the percent impervious data correlate with population density?

- How does the impervious-density relationship vary regionally (among major world environmental and cultural regions)?

- Is the impervious-density relationship consistent within countries, or within urban and rural areas of each country?

- To what extent does the impervious-density relationship vary over time? Such changes could occur through increasing urban sprawl or changes in building methods.

If there is regional consistency in the impervious-density relationship, then it will be possible to apply the results from countries with census data to neighbors with none—for example, results from Ethiopia or Kenya could be used to map population in Somalia. Temporal consistency in the impervious-density relationship will determine whether imagery can be used to map population change between censuses, though it seems unlikely that the impervious-density relationship could change radically in the space of a decade or two.

Examples

Important demographic variables—components of the "minimally essential data set" (NRC 2007, 89-96)—vary according to type of disaster and recovery stage but always include total population disaggregated by age and sex. The table below outlines other variables that are potentially useful in disaster response and are actually available for at least some countries.

Type of Event	Some Key Variables
A. War and genocide	• Baseline populations of administrative areas • Cultural variables – ethnicity, language • Household composition for estimating housing needs
B. Disease	• Populations most vulnerable to disease – children, elderly • Health status, e.g. HIV infection increases vulnerability to other diseases
C. Short-term natural disasters (hurricanes, tsunamis); nuclear disasters	• Baseline populations of small administrative areas • Vulnerable populations – elderly, lower income, migrants (especially those lacking knowledge of national language) • Housing characteristics
D. Longer-term natural disasters (climate variation, sea level rise, famine)	• Population density • Growth projection by coastal administrative unit • Ethnicity • Occupation • Urban and rural population

A. War and genocide.

Example: Genocide in Rwanda (Figure 5).

The Pop@Risk database will provide baseline populations for small administrative areas. Where available, the addition of cultural variables such as ethnic identity, language spoken, and birthplace would help predict areas with the greatest potential for ethnic conflict or genocide.
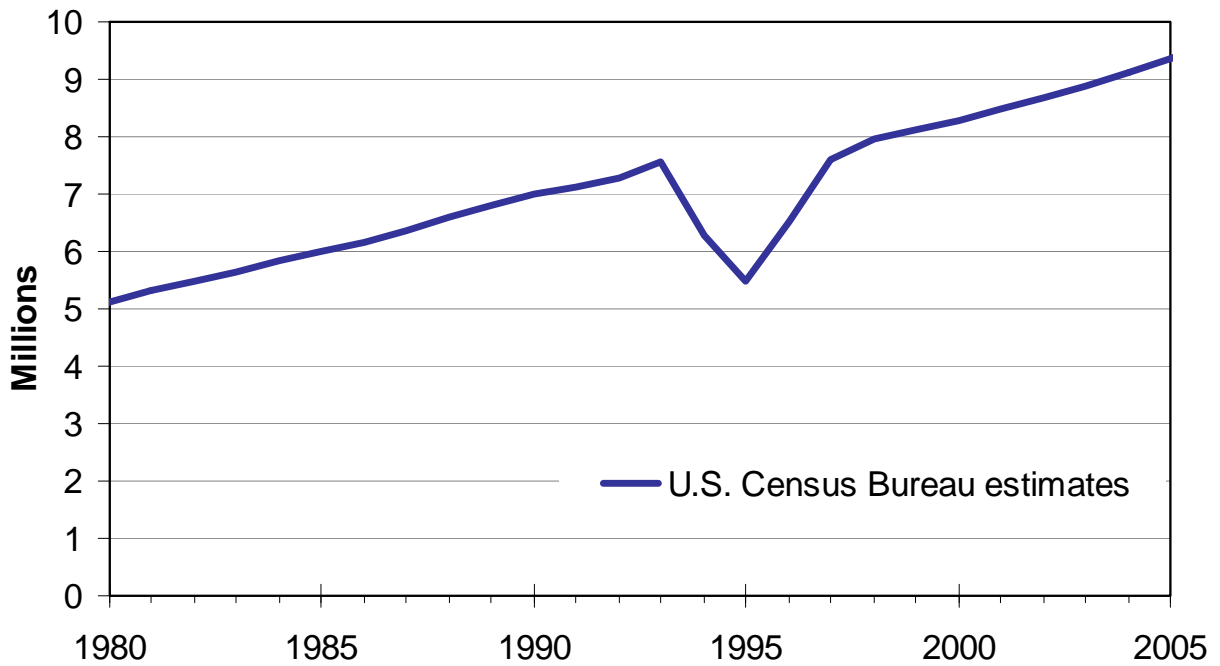

Figure 5. U.S. Census Bureau annual population estimates for Rwanda.

B. Disease.

Example: AIDS in Swaziland (Figure 6). Provisional results of the most recent Swazi census show a population considerably below UN and U.S. Census Bureau projections. Swaziland has one of the highest rates of HIV infection in the world, as reflected in the Census Bureau projection of population decline starting in the mid-2000s. However, barring a large census undercount, the effect of AIDS deaths is greater than expected.

Efficient response to a disease crisis requires information on the geographic distribution of children, the elderly, and other vulnerable groups. The HIV infection rate is itself a useful variable for a Pop@Risk database because infected people may be more vulnerable to other diseases. The ability to map vulnerable populations would be particularly important in the case of a fast-moving disease with a high fatality rate, such as an influenza pandemic.
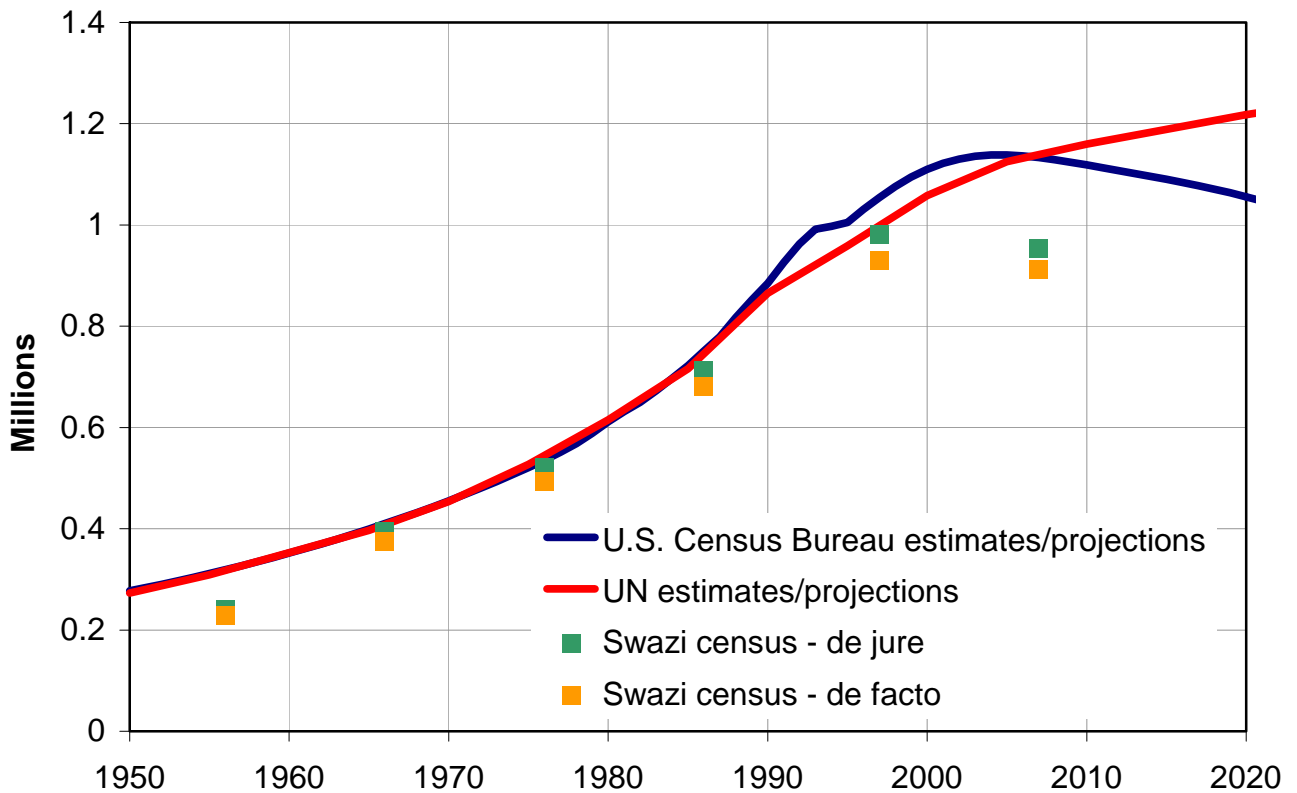

Figure 6. Census counts for Swaziland (McDermott MD 1997, Statistics South Africa 2006) along with estimates and projections from U.S. Census Bureau (2007) and UN (2007) covering years from 1950 to 2020 (the differences are due to different assumptions about life expectancy and mortality rates). The most recent Swazi census results (Swaziland Central Statistics Office 2007) are provisional.

C. Short-term disasters.

Example: Hurricane Katrina in New Orleans, USA (Figure 7).

As with a disease, a hurricane, tornado, or tsunami will differentially affect vulnerable populations. Poorer people are less able to flee and, in the longer term, rebuild their lives without assistance. Data on housing is useful for a Pop@Risk database, because different ages or types of construction are more likely to suffer damage. To the extent that the area affected by a disaster can be clearly delineated, population data for small administrative units can be used to produce overall estimates of recovery costs and to map areas where need is greatest.
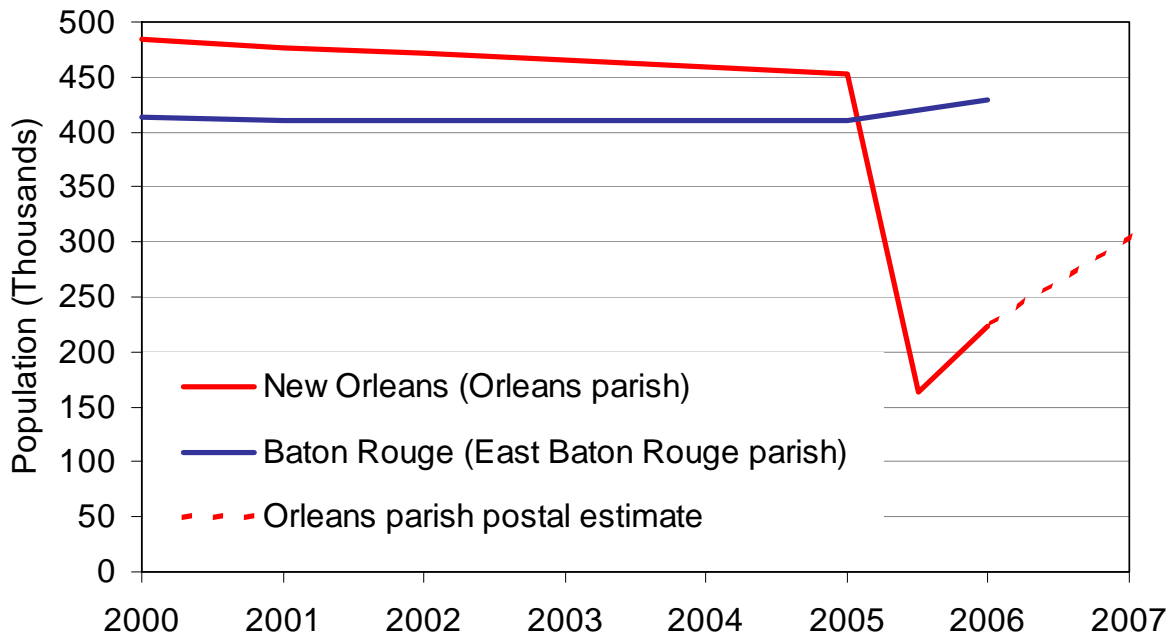

Figure 7. Population estimates before and after Hurricane Katrina (2005). Based on U.S. Census Bureau (2006), U.S. Census Bureau annual county population estimates, and postal customer counts (Greater New Orleans Data Center 2008). Baton Rouge (near New Orleans but less affected by the hurricane) was a destination for Katrina refugees.

D. Longer-term disasters.

Example: Forced migration in Ethiopia (Figure 8). Government response to climate-induced famines in the 1980s included forced migration from central Ethiopia to the sparsely populated west (Comenetz and Caviedes 2002). The different ethnic backgrounds of migrants and native residents sparked war in Gambella (Oxfam 2004), where population growth was far above the national average of 33% between 1984 and 1994.

Climate change and sea level rise have the potential to drive millions of people away from their homes and to alter ethnic patterns as people migrate. Besides ethnic data and population estimates for coastal ADMs, variables describing occupation are likely to be important for a Pop@Risk database. People who live off the land or ocean will suffer more than people with urban jobs, regardless of elevation—which will itself be an important variable for a Pop@Risk database, as shown by McGranahan *et al.*'s (2007) review of low-elevation coastal population.
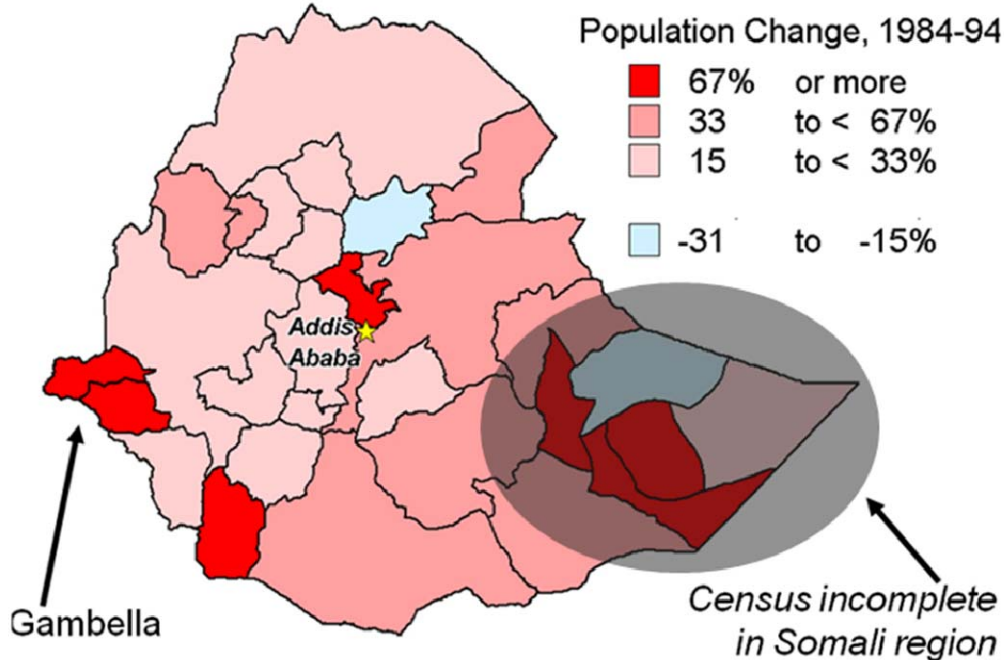

Figure 8. Population change in Ethiopia. Because provincial boundaries were redrawn (see Figure 3), the administrative units are custom "ADM1s" built from ADM3s (Comenetz and Caviedes 2002).

Acknowledgements

References

Comenetz J, Caviedes C. 2002. Climate variability, political crises, and historical population displacements in Ethiopia. *Global Environmental Change B: Environmental Hazards* 4: 113-127.

Greater New Orleans Data Center. 2008. Tracking Re-Population with Postal Data. Data onine at: www.gnocdc.org.

Homer C, Dewitz J, Fry J, Coan M, Hossain N, Larson C, Herold N, McKerrow A, VanDriel JN, Wickham J. 2007. Completion of the 2001 National Land Cover Database for the Conterminous United States. *Photogrammetric Engineering and Remote Sensing* 73: 337-341.

McDermott MD. 1997. Common Country Assessment - Swaziland, 1997 (Chapter 6: Population). Online: www.ecs.co.sz/cca (Environmental Center for Swaziland website).

McGranahan G, Balk D, Anderson B. 2007. The rising tide: assessing the risks of climate change and human settlements in low elevation coastal zones. *Environment and Urbanization* 19: 17-37.

NRC (National Research Council). 2002. *Down to Earth: Geographical Information for Sustainable Development in Africa.* Washington, DC: National Academies Press.

_____. 2007. *Tools and Methods for Estimating Populations at Risk*. Washington, DC: National Academies Press.

Oxfam America. 2004. Violence in Gambella: An Overview. Online: www.oxfamamerica.org/newsandpublications/news_updates/archive2004/art7332.html

Statistics South Africa. 2006. Country Report on Census Taking: Swaziland. Online: www.statssa.gov.za/asc/WebsiteReports/Swaziland.pdf.

Swaziland Central Statistics Office. 2007. Swaziland Population and Housing Census: Provisional Results. Online: unstats.un.org/UNSD/demographic/sources/census/2010_PHC/Swaziland/Swaziland_more.htm (United Nations Statistics Division website).

Tatem AJ, Noor AM, von Hagen C, Di Gregorio A, Hay SI. 2007. High Resolution Population Maps for Low Income Nations: Combining Land Cover and Census in East Africa. *PLoS ONE* 2(12): e1298.

UN (United Nations). 2007. World Population Prospects: The 2006 Revision Population Database. Online: esa.un.org/unpp.

_____. 2008. Census dates for all countries. Online: unstats.un.org/unsd/demographic/sources/census/censusdates.htm

U.S. Census Bureau. 2006. Special Population Estimates for Impacted Counties in the Gulf Coast Area: Hurricane Katrina. Data online at: www.census.gov/Press-Release/www/2005/katrina.htm.

_____. 2007. International Data Base (July 2007 release). Online: www.census.gov/ipc/www/idb.