

The Data Revolution: Overview and the Perspective of Demographers

Key references

“The IUSSP on a Data Revolution for Development”, *Population and Development Review* 3/2015

<http://www.undatarevolution.org/wp-content/uploads/2014/12/A-World-That-Counts2.pdf>

Data Revolution: Presentation

- Overview and Background
 - MDGs
 - SDG approach
 - Ancillary developments in the area of data
- Concerns
 - Data quality and interoperability
 - Capacity building
 - Costs and benefits (include opportunity)



Data Revolution: Overview and Background

Millennium Development Goals (MDGs)

- Core of UN development agenda 2000-2015
- 8 goals, ± 60 indicators, focusing on developing countries
- Over period, very significant improvements in many dimensions. The mobilizing effect of MDGs certainly played a role in this in some countries
- Data: task of monitoring goals spurred improvements in data collection and quality control in many countries for relevant indicators.



Data Revolution: Overview and Background

MDG indicators

- Significant data gaps, time lags in providing data and data quality problems
- Data generally inadequate for real-time monitoring of progress or as a management tool
- Examples of data gaps
 - 2013: 21 African countries had not conducted poverty survey in past 7 yrs. 2005: just 11 African countries had comparable poverty estimates.
 - CRVS (births, deaths, marriages...): only ¼ of countries in S Asia; < ½ in LA & Caribbean, and <6% in SSA have even barely acceptable systems.

Data Revolution: Overview & Background

UN Post-2015 agenda

- 2012: High-Level Panel appointed by UN Secretary-General to advise on the global development agenda after 2015, the end date for the MDGs
- The panel called for broader set of goals (SDGs) and a Data Revolution: a “new international initiative to improve the quality of statistics and information available to citizens”
- IUSSP: invited experts seminar on the role of demographers in the Data revolution (October 2014).



Data Revolution: Overview and Background

UN Post-2015 agenda: SDGs

- 17 primary goals, subdivided into targets and then indicators (100 global indicators, plus national statistics).
- Still under negotiation; goals and targets next 6 weeks.
- 3 integrated dimensions: economic development, social inclusion, and environmental sustainability. All countries.
- New focus on climate change *“which will determine whether or not we can deliver on our ambitions.”*
- Emphasis on distributional issues: *“no one is left behind.”* Data disaggregated by gender, location, income & disability, and include the very poorest and most excluded. Huge challenge for already stretched statistical systems.



Data Revolution: SDG Goals

- 1 End poverty in all its forms everywhere
- 2 End hunger, achieve food security and improved nutrition and promote sustainable agriculture
- 3 Ensure healthy lives and promote well-being for all at all ages
- 4 Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all
- 5 Achieve gender equality and empower all women and girls
- 6 Ensure availability and sustainable management of water and sanitation for all
- 7 Ensure access to affordable, reliable, sustainable and modern energy for all
- 8 Promote sustained, inclusive & sustainable economic growth, full & productive employment and decent work for all
- 9 Build resilient infrastructure, promote inclusive and sustainable industrialization and foster innovation
- 10 Reduce inequality within and among countries
- 11 Make cities and human settlements inclusive, safe, resilient and sustainable
- 12 Ensure sustainable consumption and production patterns
- 13 Take urgent action to combat climate change and its impacts*
- 14 Conserve and sustainably use the oceans, seas and marine resources for sustainable development
- 15 Protect, restore and promote sustainable use of terrestrial ecosystems, sustainably manage forests, combat
- 16 Promote peaceful and inclusive societies for sustainable development, provide access to justice for all and build
- 17 Strengthen the means of implementation and revitalize the global partnership for sustainable development



Data Revolution: Global indicators

List of global indicators that are directly demographic in nature

Goal	Indicator	Description	Standard data sources
1	7	Total Fertility Rate	Census + vital statistics + survey data
3	17	Maternal Mortality Ratio and Rate	Census + vital statistics and/or hospital data
3	18	Neonatal, Infant and Under-5 Mortality Rates	Census + vital statistics
3	19	Recommended full immunization rates	Survey data
3	20	HIV incidence, treatment & mortality	Census + vital statistics and/or hospital data
3	21	TB incidence and mortality	
3	22	Malaria incidence and mortality	
3	23	Probability of dying between 30 and 70 from any cardiovascular disease, cancer, diabetes or chronic respiratory disease	
3	25	Road traffic deaths per 100,000 population	Census + vital statistics and/or police/hospital reports
3	29	Contraceptive Prevalence Rate	Survey data
5	40	% of women aged 20-24 years old who were married or in a union before age 18	Census + vital statistics
5	44	Met demand for family planning	Survey data
16	88	Violent injuries and deaths per 100,000 population	Census + vital statistics and/or police/hospital reports

Data Revolution: Overview and Background

Key element of SDG approach = improved data

- Monitoring progress towards international and national goals
- Management tools. Data essential for decision making, policy formulation, strategic planning, resource allocation...
- Near real-time data for planning & responses to crises (Ebola...)
- Broad dissemination of data is essential for good governance: accountability and transparency.
- Needs: more variables, higher quality, available in timely and regular manner. Integrated planning: interoperability of data.
- Yet the national capacity to collect and use data varies hugely.



Data Revolution: Overview and Background

Ancillary developments /1

- G8 Open Data Charter (2013). *The Charter recognizes the central role open data can play in improving government and governance and in stimulating growth through innovation in data-driven products and services. It endorses the principle of open by default...* <http://opendatacon.org/>;
<http://opensource.com/government/13/7/open-data-charter-g8>
- Rise of civil society groups pushing for easy access to data. Government data viewed as a public good that should be available to all, to hold governments to account (transparency). Example of Civicus: <http://civicus.org/index.php/en/>



Data Revolution: Overview and Background

Ancillary developments /2

- Big data: Unstructured data collected by mobile phone firms, web-based networking sites, internet search engines, etc.
<http://www.datapopalliance.org/>; <http://www.unglobalpulse.org/about-new>; <http://chronicle.com/article/Big-Data-Big-Obstacles/151421/>;
- Other: remote sensing data (satellite), new administrative data...
- National statistical systems: no longer the sole source of data. Other providers → more competitive marketplace.
- Fast growing demand for data (firms, civil society...) → New pressures on those who collect, manage and disseminate data



Data Revolution: Overview and Background

Ancillary developments /3

- Massive improvements in computing power and in software. Ongoing efforts to render access to data more user-friendly.
- Rapidly evolving data landscape and desire to harness these developments by UN and other actors.
- For population scientists, a concern is that improvements in data quantity are no guarantee of data quality.



Data Revolution: Sense of Urgency

Ongoing rapid population growth (75m/yr) combined with rapid economic growth (3-4%/yr) → ↑↑↑ scale of world economy

Ours is a world of fabulous wealth and extreme poverty: billions of people enjoy longevity and good health unimaginable in previous generations, yet at least 1 billion people live in such abject poverty that they struggle for mere survival every day. The world economy is not only remarkably unequal but also remarkably threatening to Earth itself. The unprecedentedly large scale of the world economy is creating an unprecedented environmental crisis, one that threatens the lives and well-being of billions of people and the survival of millions of other species on the planet, and perhaps even our own. (J. Sachs (2/2015))

Data Revolution: Overview and Background

Sense of urgency /2

- Once in a generation opportunity to affect change
 - 2000-2015 MDGs. Despite flaws, goals of this sort are valuable for *mobilizing and concentrating* work.
 - SDGs: define major international initiatives for next 15 yrs
- Looming threat of climate change: need to alter rapidly production and consumption patterns to avoid irremediable harm
 - Very rapid expansion of data activities → real risks of problems (data quality...)



Data Revolution: Overview and Background

Sense of urgency /3 2015 timeline of key events

January:

- Global Conference on a Transformative Agenda for Official Statistics
- First Intergovernmental Negotiation on Post-2015 (IGN)

February: Expert Group Meeting on SDG Indicators

March: UN Statistical Commission

May: 5th Intergov Negotn on Post-2015: follow up & review processes

June -July:

- High Level Political Forum, under the auspices of ECOSOC
- Financing for Development Conference
- Final Intergovernmental Negotiation on Post-2015 (IGN)

September: **SDG Summit**

Afterwards: implementation period



Data Revolution: Overview and Background

What role for demographers?

- Population statistics underlie most indicators of human development (income per capita, literacy rates, maternal mortality, life expectancy, contraceptive use, etc.).
- Demography is perhaps the most empirical of social sciences (J. Caldwell) and demographers are preoccupied with issues of data quality
- Several classic demographic methods are of clear value for calculating relevant variables (population projections, life tables, indirect methods...) or for assessing data quality



Data Revolution:

Data quality, interoperability and gaps

Reminders

SDG indicators are (1) for monitoring progress and (2) for use as a management tool: designing and assessing programs

- Data must be precise, timely and recurrent (if possible, annual)
- Data must be disaggregated by gender, location, etc. – the principle of “leave no one behind”
- For “human development” goals, indicators are of most importance in the poorest contexts, where needs are greatest and where data is often missing and institutions tend to be weakest



Data Revolution:

Data quality, interoperability and gaps

Implications / Concerns

1. Data quality: accuracy and confidence intervals.

The aim is not just to estimate infant mortality, but to assess changes over time for specific groups.

Trends drawn from a series of point estimates may be misleading, as the trend often fall within the margin of error

Smaller populations: degree of uncertainty \uparrow \rightarrow larger samples.

Relatively rare events (e.g. maternal mortality): even with full CRVS it may be impossible to assess change for subgroups

Compromises are needed between what is desired and what is feasible: no universal “one-size-fits-all” approach can be set



Data Revolution:

Data quality, interoperability and gaps

Empirical vs. model-based estimates

- Hard data on some key variables simply do not exist in many countries (e.g., healthy life expectancy). Stop-gap estimates have been derived from complex models based on information that exists within the country and elsewhere.
- Sizable risk that model-based estimates may be considerably off or incorrectly parameterized for specific countries.
- Model estimates provide valuable information but are of little use for SDG indicators – far too uncertain. (2/2015: healthy life expectancy dropped from list of SDG indicators)



Data Revolution:

Data quality, interoperability and gaps

Empirical vs. model-based estimates

Thus:

- If at all possible, SDG indicators should be based on empirical country-level data
- This requires a major increase in data production in the poorest countries, where capacity is typically weak



Data Revolution:

Data quality, interoperability and gaps

Crucial importance of census data

- Many SDG indicators have population in their denominators
- Population size based on census data (/10 yrs). Intercensal period: population projections (robust methods) using information on components of change (migration, abortions...). Some of these are missing from the SDG indicator list.
- Census: Large (data on small groups). Used to create sampling frames for subsequent surveys. Best data available to link to other sources (e.g., handling selectivity in Big Data)
- Efforts to improve the quality and coverage of censuses should be top priority (not sufficiently stressed in UN documents).



Data Revolution:

Data quality, interoperability and gaps

New data sources

- Huge methodological difficulties in using 'new' data and in integrating them with 'old' forms
- Issues of sample selectivity at disaggregated levels, and the assessment of statistical uncertainty
- Data must be validated before use: this is common for 'old' data; unclear that equal attention is being paid to 'new' data



Data Revolution:

Data quality, interoperability and gaps

Interoperability: linking data sets together

- Methodological, legal, and privacy challenges
 - Linking: likely importance of spatial information & censuses
 - Privacy: real concerns. Tiered access...
 - Legal (jurisdictional) – e.g. Quebec in Canada
- Datasets should be accompanied by metadata about how the data were produced, edited and linked, in standardised form (e.g., the Data Documentation Initiative metadata standards for microdata in the social, behavioral and economic fields).



Data Revolution:

Data quality, interoperability and gaps

Bottom line

- Huge changes in data landscape foreseeable in upcoming years. These will transform work by quantitative social scientists, and they are needed to ensure the work is done well.
- Full potential of Big Data, interoperable data, etc. is still far off – not panaceas for the post-2015 agenda
- Combination of expanded and more regular surveys, censuses and investments in CRVS will underlie SDG work



Data Revolution: Capacity-building

- *Delivering on a Data Revolution in Sub-Saharan Africa (2014): Amanda Glassman & Alex Ezeh*
- <http://aphrc.org/wp-content/uploads/2014/07/DFAD-Policy-Brief.pdf>
-
- *“...efforts focused on collecting more data may divert attention from the underlying problems surrounding the production, analysis, and use of basic data that have inhibited progress so far. Often these problems are not merely technical but the result of implicit and explicit incentives and systemic challenges.”*



Data Revolution: Capacity-building

They identify four main obstacles

1. NSOs have limited independence and unstable budgets
 - Of the 54 countries in the African Union, just 12 have an autonomous NSO
 - Many countries: core data collection work mostly financed by external donors



Data Revolution: Capacity-building

2. Misaligned incentives leading to inaccurate administrative data
 - When data are linked to funding, local administrations have incentive to overstate successes (e.g., school enrolment)
 - Tensions between donors and governments, or different levels of government, can also lead to poor data
 - Need develop data quality checks, requiring autonomous NSOs, stable funding and competent staff



Data Revolution: Capacity-building

3. Donor priorities dominate

- Donors finance large surveys and ad hoc evaluations that fund NSOs and staff (per diems...)
- Little incentive to improve core national statistics capacity (e.g., CRVS, land registries)

4. Data dissemination

- Access to and usability of data are often severely limited
- Many NSOs/NSS are reluctant to disseminate data and/or lack the capacity to publish and broadly diffuse data according to international standards



Data Revolution: Capacity-building

Another problem: human resources

- Essential to have enough qualified staff to design and direct data collection, clean and disseminate the data, and who are capable of integrating new approaches and software into NSO work
- Serious dearth in many NSOs, due to inadequate training and unstable or low salaries (turnover)
- Need viable well-funded institutions and also time to expand training and rethink training curriculums (e.g., integrating demographic methods and theories in to social statistics)



Data Revolution: Capacity-building

Without effective institutions and competent staff

- DR initiatives risk bypassing rather than reinforcing national capacity
- Data risks being of poor quality
- But building institutions and training staff takes time...

Some institutions are working on these issues (e.g., Paris21).

But the post-2015 Data revolution is mostly focused on rapidly producing data for SDGs and seems mostly to lightly pass over such issues. Recognition of weak NSOs → second-best solutions (e.g., less frequent surveys and interpolation).



Data Revolution: Overview and Background

Costs of data for the SDG indicators

- MDGs: data costs estimated at 28b USD /15 yrs, and the data were far from adequate
- Jervan (9/2014) estimated cost of 169 indicators at 254b USD → ± 150b for 100 (\$10b /year). His analysis neglected new technologies (\$↓) as well as capacity-building expenses, expanding CRVS... (\$↑).
- UN SDSN et al. estimate costs at 1b USD/year, half from national sources. Does not seem to include current ongoing costs (censuses, DHS...) nor major capacity-building efforts in the poorest countries



Data Revolution: Overview and Background

- Opportunity costs: Huge emphasis on SDGs mobilizes staff and resources, and thus effectively deemphasizes other data collection efforts (including of causal determinants of SDGs) or activities
- Costs versus benefits: Yet what are the costs of not having this information which could underlie better and more efficient policies and programs?
- Decisions: Financing for Development conference in Addis Ababa, 7/2015
<http://www.un.org/esa/ffd/overview/third-conference-ffd.html>



Data Revolution: Capacity-building

Haishan Fu* has said that developing the capacity of poorer countries in the area of data (collection, analysis & diffusion) should be a major goal in and of itself.

- Recognition that this is a long-haul process
- SDG agenda: Extremely ambitious but feels very rushed
- Insufficient attention paid to some key data quality issues
- Opportunity costs and political process (what emphasized, what is not)
- Institution and staff capacity-building needs are overly neglected in the poorest countries

* Director, Development Economics Data Group, World Bank. PhD in demography

