

Stuffing data gaps with dollars: What will it cost to close the data deficit in poor countries?

Laurence Chandy and Christine Zhang

August 2015

ABSTRACT

We calculate a financing gap for data needs based on an analysis of existing survey coverage for three pillars of development data: the population census, living standard surveys, and health surveys. We find that, at least in the case of living standard surveys and health surveys, the global financing gap is trivially small—approximately \$23 million a year. This points to other constraints beyond finance that stand in the way of complete data coverage. We conclude by recommending three immediate actions the IMF, World Bank, USAID, and UNICEF could take to advance this agenda.

Last month, Addis Ababa, Ethiopia played host to a major United Nations conference on financing for development. Among the issues that featured prominently was financing for data. This reflects an emerging consensus that data can play a powerful enabling role in the development of poor countries, but that existing efforts to collect data are under-resourced.

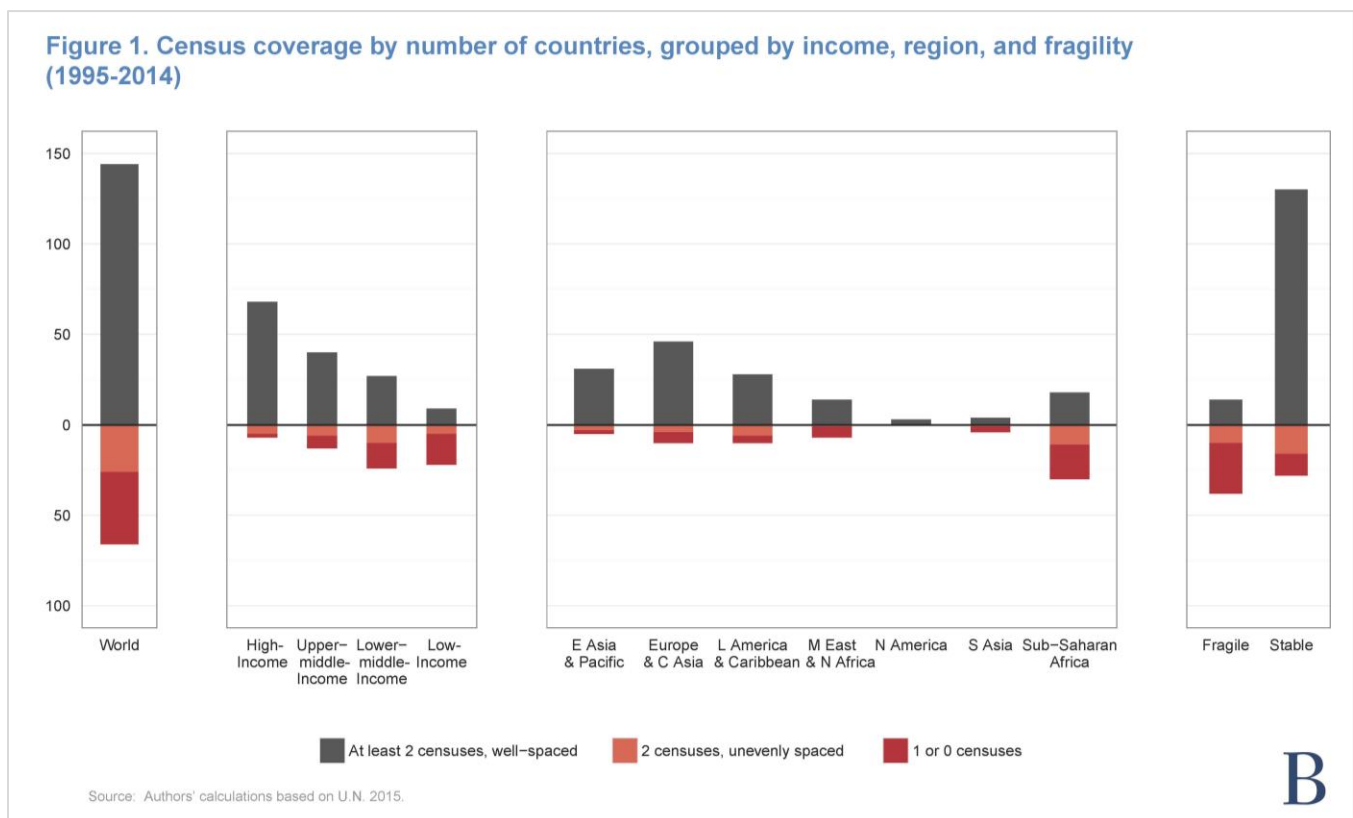
A series of papers published over the past year have attempted to calculate a financing gap for data needs. Since costing development outcomes is an intrinsically difficult exercise, they narrow their focus on the more tangible cost of expanding data collection through surveys. The [first, by Morten Jerven](#), made a splash with the claim that monitoring the post-2015 Sustainable Development Goals globally would cost a jaw-dropping \$254 billion, up from an estimated (but never fully delivered) \$27 billion for data collection under the Millennium Development Goals. [Gabriel Demombynes and Justin Sandefur](#) pared this down by focusing on the poorest countries whose data needs are most likely to be supported by donors and by limiting data collection to a basic package of surveys. They estimate a financing gap of \$300 million a year to support countries where annual per capita income falls under \$2,000. Finally, the [U.N.'s Sustainable Development Solutions Network](#) (SDSN) attempted some more detailed calculations that net out existing budgetary commitments from governments and donors. Its estimate of the financing gap was just \$100-\$200 million per year.

Oddly missing from all three studies is an assessment of the degree to which existing survey coverage is incomplete. A gap analysis of surveys is surely the most logical basis for determining the gap in survey financing. We use this approach to estimate the financing gap for three pillars of development data: the population census, living standards surveys, and health surveys. Along the way we examine the characteristics of countries with the most serious data gaps, the suitability of existing standards for data needs, and the challenges and opportunities entailed in boosting financing for data.

Population censuses

Countries are expected to undertake a population census once every 10 years in line with standards codified in the International Monetary Fund’s (IMF) General Data Dissemination System (GDDS). The U.N. Statistics Division plays a coordinating role by defining each census round and encouraging all member countries to participate.

To assess strict compliance with the GDDS standard requires an assessment of participation over a 20-year period. We find that 66 countries currently fail to meet the standard (Figure 1).¹ However, 26 of these fell just short by allowing the number of years between each census to extend narrowly beyond a decade. This points to the logistical complexities of successfully executing a census on schedule and on budget. For these countries, implementation capacity as opposed to financing capacity would appear to be the more relevant constraint.



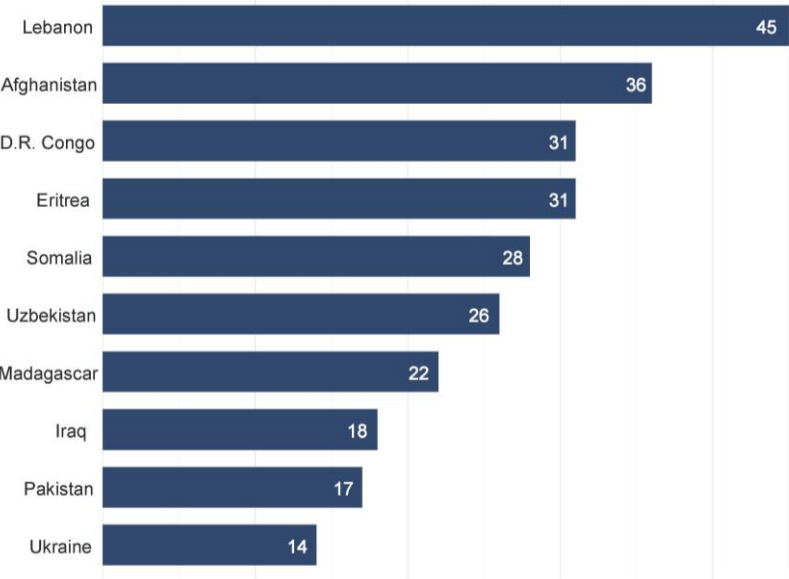
The remaining 40 countries unambiguously fall foul of the standard. These are disproportionately low-income (more than half of all low-income countries are found among the 40) indicating that finance may be an important factor explaining non-compliance. However, they also make up more than half of the OECD’s list of fragile states, which points to other possible underlying challenges: instability, or a

¹ Census data is drawn from the [U.N. Statistics Division website](http://www.unstats.un.org/). Countries are judged to be GDDS compliant if they have undertaken two censuses over this period separated by no more than 10 years, or if they have comprehensive administrative registries capable of generating the same data.

lack of will or capacity by the state to undertake data collection. Figure 2 shows the laggards who have gone the longest without carrying out a new census: Lebanon tops the list at 45 years.

Participation in each census round is improving over time. 31 countries were missing from the 2000 round whereas only 21 were excluded in the 2010 round. Assuming for simplicity's sake that financing is the sole constraint to complete census coverage, we calculate the size of the financing gap during the 2010 round by applying SDSN estimates of census costs per capita (around \$2 per head) to these 21 countries. This generates a price tag of just over a billion dollars, or **\$108 million a year** over 10 years. By design, this cost is driven up by the largest countries missing from the 2010 round. Pakistan and DR Congo can account for nearly half of the estimated 500 million people excluded from this census round—and therefore half the financing gap.²

Figure 2. Years since last census for laggard countries



Source: Authors' calculations based on U.N. 2015, World Bank 2015.

B

Living standards surveys

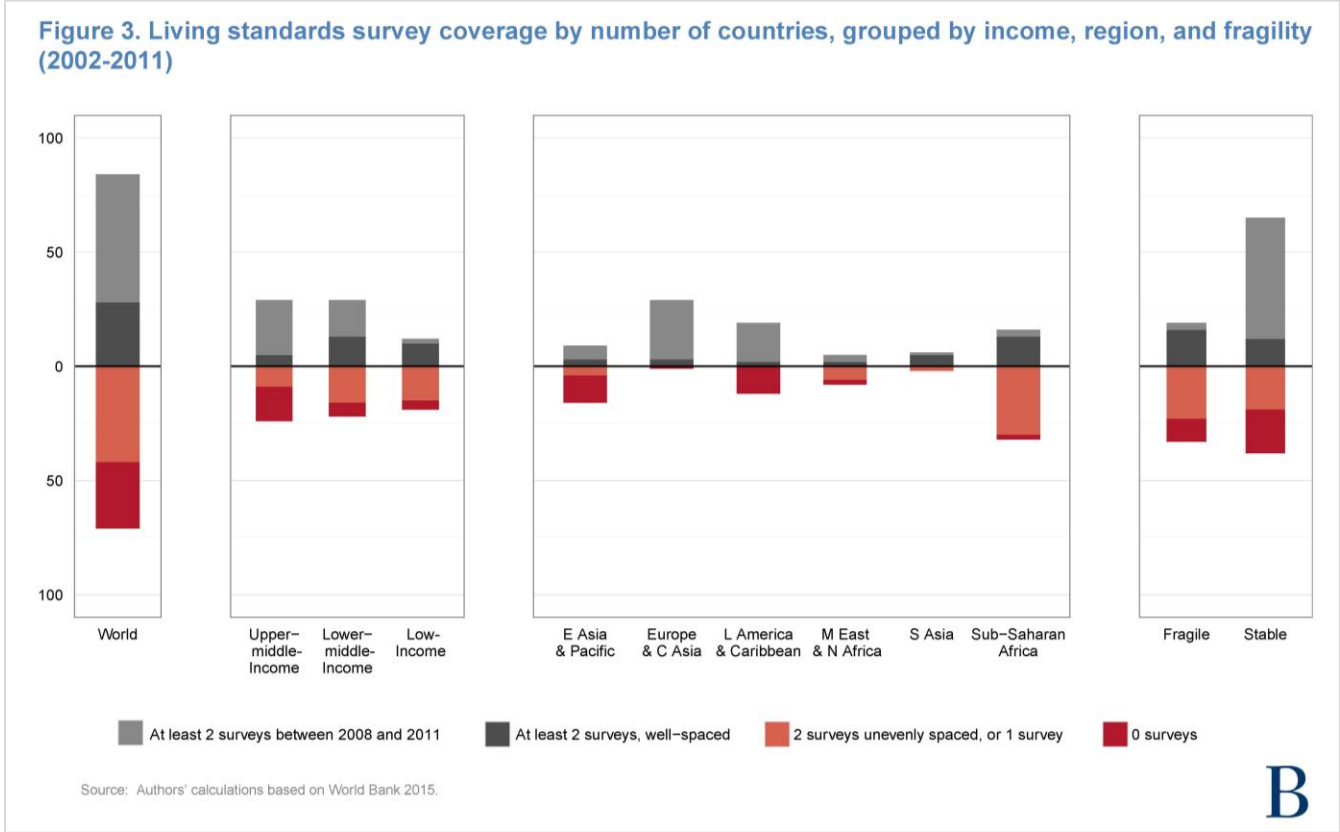
Living standards surveys, often referred to as income or consumption surveys, are the leading source of poverty and inequality measures. The GDDS recommends that countries undertake these surveys at least every 5 years.³ The World Bank and regional development banks support many developing countries in survey design, implementation, and analysis.

A [recent study by Umar Serajuddin and World Bank colleagues](#) assessed compliance with the GDDS standard among developing countries. It found that half of the 155 countries studied fall short of the standard.

² Pakistan's National Socio-Economic Registry, run by the Benazir Income Support Program, contains details on the vast bulk of the country's population (98 percent of all adults), but it is not considered a complete substitute for a census. Guatemala is among the group of 21 countries and its Registro Unico de Usuarios Nacional performs a similar role.

³ The GDDS standard is officially given as "at least every 3 to 5 years"—a sublime example of statistical illiteracy.

Our replication of their analysis confirms these results (Figure 3).⁴ Despite a steady improvement in the rate of compliance since the start of the century more than a billion people live in countries with insufficient data on living standards—approximately the same number of people who are believed to live in extreme poverty around the world. Twenty-nine countries do not have a single reported survey of living standards since the year 2000, although more than half of these have a population under 200,000 people, highlighting the capacity challenges—both financial and operational—of delivering on the post-2015 mantra to “leave no-one behind.”⁵



Small population size is not the only correlate of data deficiency. Fragile states again feature prominently among the 71 countries that do not comply with the GDDS standard. Also notable is the large regional variation in compliance. This suggests a potential role for regional coordinating mechanisms and institutions to raise the standard of data monitoring.

Assuming again that insufficient financing can explain all existing data gaps and relying on SDSN cost estimates of \$1.7 million per living standards survey, we estimate the global cost of achieving universal GDDS compliance is a paltry **\$17 million a year** (i.e., 10 additional surveys per year) based on existing survey coverage. Living standards surveys in developing countries are often subsidized by foreign assistance, so this could be realized with only a modest reallocation of donor funds—\$17 million a year

⁴ Surveys of living standards are drawn from PovcalNet and World Development Indicators. Our findings differ marginally from those of Serajuddin et al., likely due to minor updates in the World Development Indicators.
⁵ Our analysis by design excludes surveys whose results were not made accessible to the World Bank or were rejected by the Bank due to quality concerns—see later discussion on implications for data financing.

is approximately what the World Bank spends on tiger preservation. Indeed, this sum is so small that it makes our assumption that finance is the binding constraint to GDDS compliance seem almost self-defeating.

In any case, it is debatable whether a survey of living standards every 5 years is an appropriate benchmark to satisfy data needs in today's developing countries. This benchmark seems better suited for providing a historical record of well-being than for informing policy, especially given the growing adoption of large-scale targeted poverty programs. At present, 26 developing countries conduct surveys of living standards yearly and a further 23 administer them every 2 years, demonstrating the feasibility of more regular assessments. If the GDDS standard were raised so that all developing countries were expected to conduct a living standards survey every 2 years, the financing gap would rise to \$66 million a year. Were this standard to be limited to countries with a population of over 1 million and the cost restricted only to subsidizing low-income and lower-middle income countries, \$35 million of additional expenditure a year would be sufficient to close the gap.

Health surveys

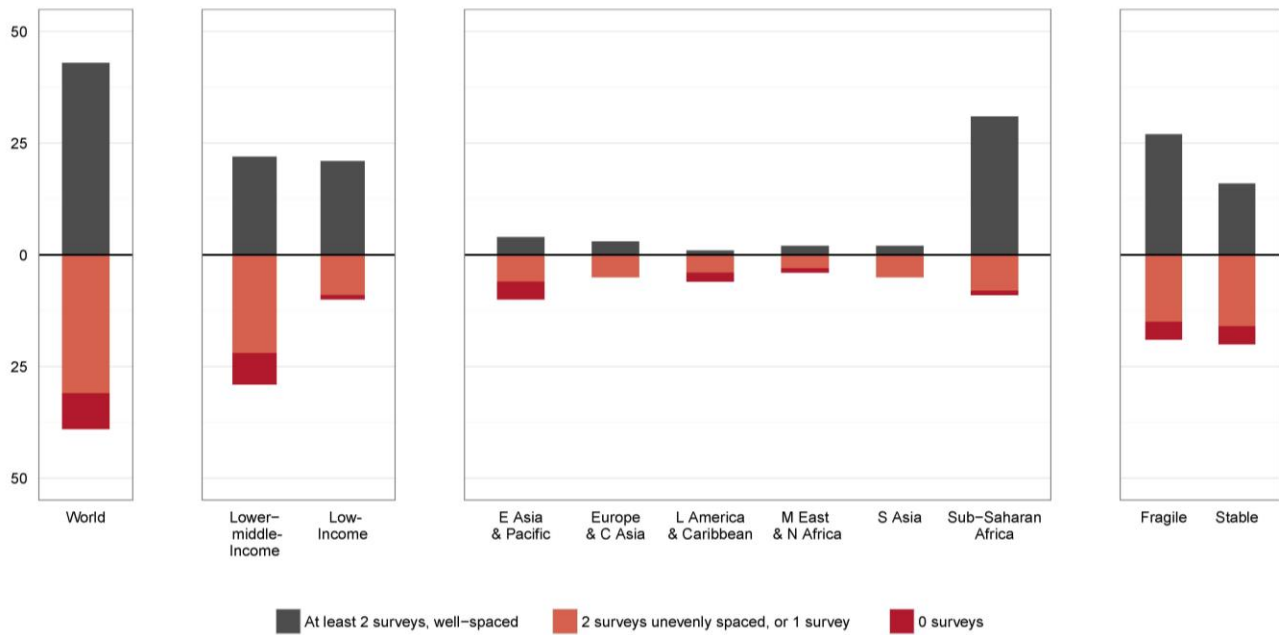
Whereas censuses and living standard surveys are employed in all developed and developing countries, health surveys only play a role in the latter. They serve as a substitute for effective administrative data, specifically civil registration and vital statistics systems and hospital records, which are typically weak or missing in poor countries. Among the three common standardized health surveys, two are designed and sponsored by donors: the Demographic and Health Survey (DHS) by USAID and the Multiple Indicator Cluster Survey (MICS) by UNICEF. (The Pan Arab Project for Family Health, or PAFAM, is the other.)

An analysis of the coverage of the three surveys in developing countries suggests that reliance on health surveys drops off—and thus administrative data picks up—for countries with per capita income over \$7,000 a year.⁶ We therefore focus on low-income and lower-middle-income countries (where annual per capita income is below \$4,125), which we can reasonably assume depend on health surveys to generate reliable measures of health outcomes.

The GDDS stipulates that countries generate health data on an annual basis. However, it does not specify how often health surveys ought to be undertaken when administrative data is lacking. In the absence of an independent benchmark of good health survey coverage, we apply the GDDS standard for living standards surveys of at least every 5 years (Figure 4). Of 82 low and lower-middle-income countries, 39, home to 1.9 billion people, fail to meet this standard over the past decade based on a review of DHS, MICS, and PAFAM surveys. Eight countries have no reported health survey at all. In contrast to census and living standard surveys, there is no improvement in coverage over the past decade by this measure (Figure 5).

⁶ Health surveys are drawn from the DHS, MICS, and PAFAM websites. Standard, continuous, and interim DHS surveys are included. Sub-national surveys and surveys that are not nationally representative are excluded.

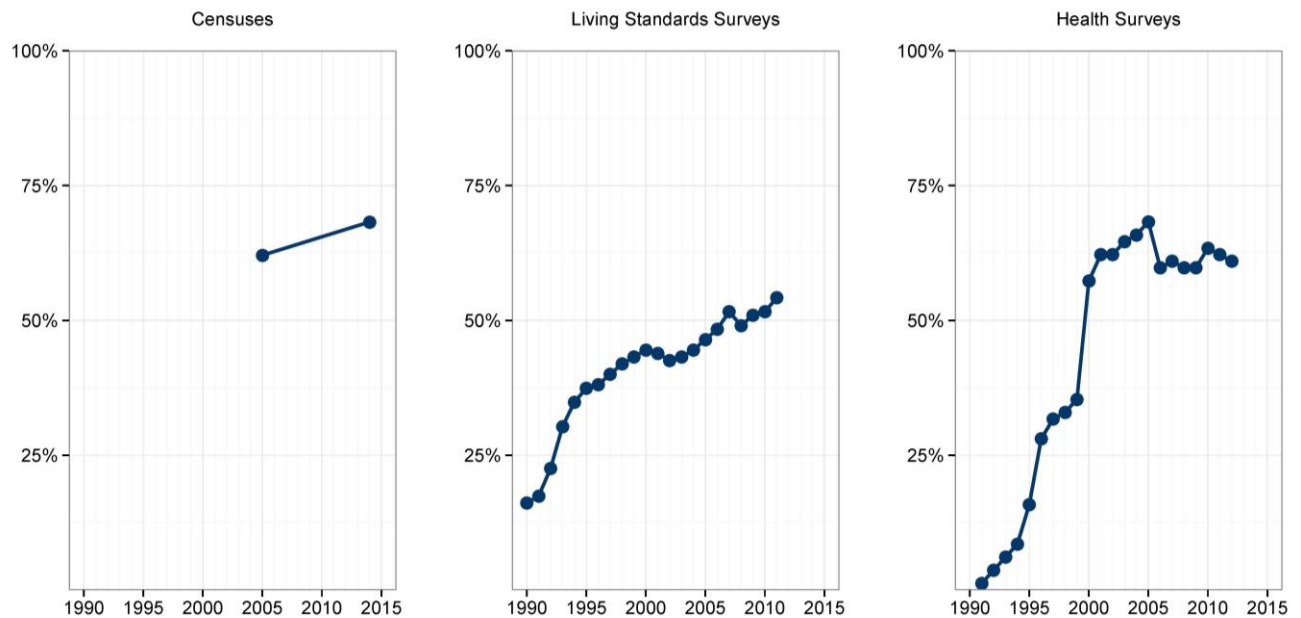
Figure 4. Health survey coverage by number of countries, grouped by income, region, and fragility (2005-2014)



Source: Authors' calculations based on MEASURE DHS and ICF International 2015, PAPPAM 2015, UNICEF MICS 2015.

B

Figure 5. Share of countries with satisfactory data coverage over time



Note: Comparisons across the charts should be made with care given the differences in the countries under assessment. For censuses, analysis is based on adherence to GDDS standard over the preceding twenty years, each capturing two census rounds, for 210 countries. For living standards surveys, analysis is based on adherence to GDDS standard over the preceding ten years, for 155 developing countries. For health surveys, analysis is based on adherence to GDDS-equivalent standard over preceding ten years, for 82 low-income and lower-middle-income countries. In addition to DHS, MICS and PAPPAM surveys, World Health Surveys conducted in 70 countries between 2002 and 2004 by the World Health Organization are included for this analysis.

Source: Authors' calculations based on MEASURE DHS and ICF International 2015, PAPPAM 2015, U.N. 2015, UNICEF MICS 2015, WHO 2015, World Bank 2015.

B

The composition of the 39 countries with good health survey coverage is striking. Low-income countries are more likely to have good survey coverage than lower-middle-income countries. Among low and lower-middle-income countries, fragile states perform better than stable ones. And countries in sub-Saharan Africa have greater coverage than those in all other regions. This can be explained by the dominant role USAID and UNICEF play in commissioning health surveys. For instance, DHS surveys are only carried out in least developed countries and countries receiving U.S. foreign assistance. In most cases, the surveys are carried out at the request of the local USAID mission.

To estimate the financing gap for health surveys, we use the average of SDSN's cost estimates for a DHS and MICS survey, which is \$1.3 million. Applying this to the gaps in existing health survey coverage indicates a financing gap of **\$6 million a year**. Again, this figure seems too small to justify being called a financing gap.

Implications for data financing

With growing excitement about the potential for sparking a data revolution in the developing world, a focus on traditional data collection through surveys may seem anachronistic. Yet we believe surveys will continue to play a foundational role in building knowledge around the state of developing economies. (This is especially true when it comes to harnessing the potential of big data, which hinges on a plentiful supply of traditional data to assess what development indicators big data can replace or predict, and their degree of accuracy.) That makes raising the level of survey coverage a data priority.

At the Addis conference, a Global Partnership for Sustainable Development Data was announced that promises to mobilize new resources for data collection. However, when it comes to improving survey coverage, it is not clear that finance is the binding constraint. Weak capacity, instability and bad governance at the country level, combined with coordination failures internationally, seem equally if not more relevant.

In poor countries finance often remains an important obstacle to administering more surveys. But it is a mistake to assume that national financing gaps necessarily translate into a global financing constraint. This may help to explain why [a recent World Bank evaluation](#) found that financing is ranked much higher on a list of constraints to data collection by developing country governments than by Bank staff: each is talking about different things.

An analysis of survey coverage necessarily leads to a focus on survey quantity over quality. Yet poor data quality is an equally important problem for many developing countries. Furthermore, the issues of quantity and quality are linked. Incomplete survey coverage can be explained not just by the absence of surveys but by surveys whose results have to be rejected due to quality concerns. A recent update of the World Development Indicators resulted in [66 of 478 new poverty estimates being dropped](#). Improving the quality of data collected through surveys likely has cost implications even if these are less easy to quantify.

New efforts to boost investment in data collection and statistical capacity following from the Addis conference should think seriously about how future approaches to data development can learn from

previous experience. Living standards and health surveys represent two distinct models for data development. Living standards surveys tend to be country-led, providing a more likely path towards sustainable and improved statistical capacity. Health surveys tend to be donor-driven, resulting in better coverage among the poorest countries; greater transparency of results (including free access to micro data); and survey designs that are better harmonized, generating more comparable results over time and between countries. These trade-offs are unfortunate and avoidable.

Recommendations

The above analysis indicates that addressing data gaps is less straightforward than it seems and requires more than simply stuffing them with money. However, we identify three easy wins for the donor community to advance this agenda:

1. The IMF should review the GDDS standard for the frequency of living standard surveys with a view to raising the standard for all but the smallest countries to better serve data needs.
2. The World Bank should commit to supporting all developing countries that do not adhere to the GDDS standard on living standard surveys so that full survey coverage is achieved by 2020. This support should be made conditional on survey documentation and results being granted public access.
3. USAID and UNICEF should, through a coordinated effort, commit to supporting all low-income and lower-middle-income countries that do not have a health survey at least every 5 years so that full data coverage is achieved by 2020. Simultaneous investments should be made in these countries to strengthen civil registration and vital statistics systems and hospital records as part of a longer term solution to data capacity.